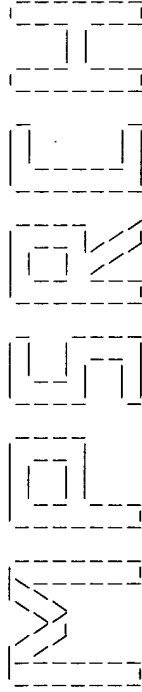


HUFF

09(081707

SEQ ID: 7, 10, 11, 12



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:19:27 1999; MasPar time 3.25 Seconds
Tabular output not generated. 39.202 Million cell updates/sec

Title: >US-09-081-707-7
Description: (1-6) from US09081707.pep
Perfect Score: 39
Sequence: 1 HSSKLO 6
Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 12.999; Variance 29.123; scale 0.446

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	Description	Pred. No.
1	36	92.3	992 34	W69742	1.78e+02
2	36	92.3	992 34	W69740	1.78e+02
3	35	89.7	742 14	R74094	2.46e+02
4	34	87.2	724	W27419	3.38e+02
5	34	87.2	118 18	R97559	3.38e+02
6	34	87.2	118 20	W08364	3.38e+02
7	34	87.2	127 24	W21854	3.38e+02
8	34	87.2	127 24	W21855	3.38e+02
9	34	87.2	127 24	W21844	3.38e+02
10	34	87.2	127 24	W21853	3.38e+02
11	34	87.2	127 24	W21848	3.38e+02
12	34	87.2	127 24	W21852	3.38e+02
13	34	87.2	179 6	R10310	3.38e+02
14	34	87.2	188 6	R30779	3.38e+02
15	34	87.2	3084 35	W50891	3.38e+02
16	33	84.6	15 28	W49219	4.63e+02

17	33	84.6	210 35	W70284	Anaplasma marginale m	4.63e+02
18	33	84.6	293 27	W35351	Haematozoon pluvial	4.63e+02
19	33	84.6	305 27	W35350	Haematozoon pluvial	4.63e+02
20	33	84.6	406 29	W40035	Novel human protein d	4.63e+02
21	33	84.6	407 38	W73420	Human secreted protei	4.63e+02
22	33	84.6	507 13	R66631	HCV J1 NS3-NS4 domain	4.63e+02
23	33	84.6	532 35	W49700	Human flavin-containi	4.63e+02
24	33	84.6	680 2	P94508	Sequence encoded by p	4.63e+02
25	33	84.6	739 8	R41334	91 kD ISGF-3alpha	4.63e+02
26	33	84.6	749 13	R72080	Mouse Stat1 (Stat91)	4.63e+02
27	33	84.6	749 18	W03172	Mouse STAT1	4.63e+02
28	33	84.6	750 35	W62994	Human Stat1-alpha pro	4.63e+02
29	33	84.6	750 18	W03168	Human Stat1-alpha	4.63e+02
30	33	84.6	750 13	R72078	Human Stat91	4.63e+02
31	33	84.6	839 3	R15629	Capsid region of cyno	4.63e+02
32	33	84.6	841 2	P94511	Sequence expressed by	4.63e+02
33	33	84.6	3084 2	P94758	Sequence of mouse lam	4.63e+02
34	32	82.1	212 2	R06893	Tilapia prolactin I	6.32e+02
35	32	82.1	212 2	R08121	Modified tilapia prol	6.32e+02
36	32	82.1	282 16	R48731	G-protein coupled hum	6.32e+02
37	32	82.1	282 19	W02703	G-protein coupled hum	6.32e+02
38	32	82.1	325 2	P70428	Polyptide encoded b	6.32e+02
39	32	82.1	376 17	R90541	pJG4-5-CDK-BP clone #	6.32e+02
40	32	82.1	456 3	R14144	Tyrosine phenol-lyase	6.32e+02
41	32	82.1	462 17	R89583	Human semenogelin I	6.32e+02
42	32	82.1	462 26	W33420	Human semenogelin I	6.32e+02
43	32	82.1	729 17	R89275	Yeast coagulation pro	6.32e+02
44	32	82.1	1021 24	W23281	Allorreaction associat	6.32e+02
45	31	79.5	79 16	R78633	Partial MAP kinase ph	8.58e+02

ALIGNMENTS

RESULT 1

ID W69742 standard; protein; 992 AA.
AC W69742;
DT 26-OCT-1998 (first entry)
DE SAPAP2 protein.
KW Human; SAPAP2; SAPAP1; animal protein; PSD-95/SAP90; diagnosis;
KW nervous disease; functional interference; structural interference;
OS membrane associated guanylate kinase; neuronal disease.
QS Homo sapiens.
PN IL0201478-A.
PD 04-AUG-1998 011715.
PF 24-JAN-1997, 011715.
PR 24-JAN-1997; JP-011715.
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (TAKE/) TAKEUCHI M.
DR WPI: 98-474492/41.
PT DNA encoding new animal protein SAPAP 2 - useful for diagnosis and
PT treatment of nervous system diseases
PS Claim 1; Page 7-9; l1pp; Japanese.
CC The present invention is: (A) an animal protein having an amino acid
CC the present invention is: (A) an animal protein having an amino acid
CC sequence substantially the same as SAPAP2; (B) a cDNA sequence encoding
CC the amino acid sequence of SAPAP2 or (C) an amino acid sequence
CC substantially same as SAPAP2; and (D) a genomic DNA sequence hybridised
CC by the above cDNA or its partial sequence. SAPAP2 is a new animal
CC protein which combines specifically with PSD-95/SAP90 and its related
CC protein and is useful for the diagnosis, prevention and treatment of
CC various nervous diseases caused by functional or structural interference
CC of nervous system.
SQ Sequence 992 AA;

Query Match 92.3%; Score 36; DB 34; Length 992;
Best Local Similarity 83.3%; Pred. No. 1.78e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 402 hspklq 407
|||
QY 1 HSSKLO 6

```
RESULT 2
ID W59740 standard; protein; 992 AA.
AC W59740;
DT 26-OCT-1998 (first entry)
DE SAPAP1 protein.
KW Human; SAPAP1; SAPAP2; animal protein; PSD-95/SAP90; diagnosis;
KW nervous disease; functional interference; structural interference;
KW membrane associated guanylate kinase; neuronal disease.
OS Homo sapiens.
PN J10201477-A.
PD 04-AUG-1998.
PF 24-JAN-1997; 011714.
PR 24-JAN-1997; JP-011714.
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (TAKEU) TAKEUCHI M.
DR WPI: 98-474491/41.
PT New protein SAPAP1 - used for, e.g. diagnosis and prevention of
PT various neuronal diseases
PS Claim 1: Page 4-7; 12pp; Japanese.
CC The present sequence represents the SAPAP1 protein having a 992 amino
CC acid (aa) sequence. Also described in the present invention are: (1) an
CC animal protein having an aa sequence substantially homologous to SAPAP1;
CC (2) cDNA sequence encoding SAPAP1, or an aa sequence substantially
CC homologous to SAPAP1, and (3) a genomic DNA sequence hybridised to the
CC cDNA or its partial sequence. SAPAP1 is a novel animal protein specific
CC for PSD-95/SAP90 and its related protein, and may be useful for the
CC diagnosis, prevention and treatment of various neuronal diseases caused
CC by functional or structural interference of nervous system.
SQ Sequence 992 AA;

Query Match 92.3%; Score 36; DB 34; Length 992;
Best Local Similarity 83.3%; Pred. No. 1.78e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 402 hspklq 407
QY 1 HSSKQ 6
||:|:|

RESULT 3
ID R74094 standard; Protein; 742 AA.
AC R74094;
DT 04-JAN-1996 (first entry)
DE Human zona pellucida-2 (H2P-2) protein.
KW Human zona pellucida; H2P-2; protein; vaccine; antigen;
KW contraceptive; ovary.
OS Homo sapiens.
PN J07099974-A.
PD 18-APR-1995.
PF 05-OCT-1993; 249404.
PR (TOFU) TONEN CORP.
PA (TOFU) TONEN CORP.
DR WPI: 95-182067/24.
DR N-PSDB: Q92254.
PT A DNA coding human zona pellucida-2 protein - used as a vaccine
PT antigen
PS Claim 1: Page 8-11; 11pp; Japanese.
CC The human zona pellucida-2 (H2P-2) protein may be used as a vaccine
CC antigen. It can be artificially synthesised using recombinant
CC techniques. Partial sequences of the H2P-2 gene were subcloned, the
CC sequences and primers used in cloning and subcloning are given in
CC Q92255-69.
SQ Sequence 742 AA;

Query Match 89.7%; Score 35; DB 14; Length 742;
Best Local Similarity 66.7%; Pred. No. 2.46e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 472 hatklq 477
QY 1 HSSKQ 6
||:|:|
```

```
RESULT 4
ID W27419 standard; peptide; 7 AA.
AC W27419;
DT 19-DEC-1997 (first entry)
DE CDR2 from light chain variable region of KM1259 antibody.
KW Complementarity determining region; CDR; light chain; treatment;
KW variable region; murine; mouse; human; interleukin 5; IL-5;
KW receptor; alpha chain; monoclonal antibody; hybridoma; detection;
KW assay; diagnosis; allergic respiratory disease; chronic bronchitis.
OS Mus spp.
PN W09710354-A1.
PD 20-MAR-1997.
PF 11-SEP-1996; J02588.
PR 11-SEP-1995; JP-232384.
PA (KYOW) KYOWA HAKKO KOGYO KK.
PI Aizawa H, Furuwa A, Hanai N, Iida A, Koike M;
PI Nakamura K, Takatsu K;
DR WPI: 97-202249/18.
PT Antibody against alpha-chain of human interleukin 5 receptor --
PT useful for diagnosis and treatment of respiratory allergic diseases,
PT e.g. chronic bronchitis
PS Claim 8: Page 165; 238pp; Japanese.
CC The present sequence is complementarity determining region 2 (CDR2)
CC from the light chain variable region of the murine anti-human
CC interleukin 5 receptor alpha chain (hIL-5R alpha) monoclonal
CC antibody (MAB) KM1259. KM1259 is produced by the hybridoma
CC FERM BP-5134, which was prepared by immunising Balb/c mice with
CC hIL-5R alpha, fusing spleen cells obtained from the mice with mouse
CC myeloma P3-U1 cells and screening the resultant hybridomas. The MAB
CC can be used to detect or assay for hIL-5R alpha and cells
CC expressing it on their surface, especially to diagnose allergic
CC respiratory diseases, e.g. chronic bronchitis. It can also be used
CC to treat such diseases.
SQ Sequence 7 AA;

Query Match 87.2%; Score 34; DB 24; Length 7;
Best Local Similarity 66.7%; Pred. No. 3.38e-02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 htsrlq 6
QY 1 HSSKQ 6
||:|:|

RESULT 5
ID R97559 standard; Protein; 118 AA.
AC R97559;
DT 14-SEP-1996 (first entry)
DE Tomato p2130 cDNA-encoded protein.
KW Ovary; ovule; fruit; tomato; cotton; melanin; vector;
KW transgenic plant.
OS Lycopersicon esculentum cv. UC82B.
PN US5530185-A.
PD 25-JUN-1996.
PF 19-JUL-1989; 382518.
PR 19-JUL-1989; US-382518.
PR 17-JUL-1990; US-554195.
PR 29-DEC-1992; US-998158.
PA (CALT) CALGENE INC.
PI Martineau BM, Reilley AA, Stalker DM;
DR WPI: 96-308822/31.
DR N-PSDB: T31823.
PT DNA construct for expressing melanin synthesis gene in plant ovule
PT cells - contains promoter from the tomato p2130 gene, also binary
PT vector and transgenic plants, esp. cotton, contg. construct
PS Example 3; Fig. 1A-B; 25pp; English.
CC The tomato p2130 clone contains a 564 bp insert of cDNA (T31823)
CC detected only in a tomato cDNA library prepd. from pre-anthesis
CC RNA. It codes for a protein (R97559) of unknown function. The
CC clone was isolated from a library prepd. from cDNA of pre-anthesis
CC stage ovaries. The p2130 transcriptional initiation region is
CC considered to be ovary-specific and can be utilised in DNA
CC constructs for the expression of heterologous genes, partic. in
```


CC Balb/c mice with hIL-5R alpha, fusing spleen cells obtained from
 CC the mice with mouse myeloma P3-U1 cells and screening the resultant
 CC hybridomas. The MAb can be used to detect or assay for hIL-5R alpha
 CC and cells expressing it on their surface, especially to diagnose
 CC allergic respiratory diseases, e.g. chronic bronchitis. It can also
 CC be used to treat such diseases.
 SQ Sequence 127 AA;

Query Match 87.2%; Score 34; DB 24; Length 127;
 Best Local Similarity 66.7%; Pred. No. 3.38e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 70 htsrlq 75
 1:1:11
 QY 1 HSSKLQ 6

RESULT 9
 ID W21844 standard; Protein; 127 AA.
 AC W21844;
 DT 19-DEC-1997 (first entry)
 DE Light chain variable region of KM1259 antibody.
 KW Complementarity determining region; CDR; light chain; treatment;
 KW variable region; murine; mouse; human; interleukin 5; IL-5;
 KW receptor; alpha chain; monoclonal antibody; hybridoma; detection;
 KW assay; diagnosis; allergic respiratory disease;
 KW chronic bronchitis.
 OS Mus spp.
 FH Key Location/Qualifiers
 FT peptide 1..20
 FT /label= sig_peptide
 FT peptide 21..127
 FT /label= mat_peptide
 FT region 44..54
 FT /label= complementarity_determining_region_1
 FT region 70..76
 FT /label= complementarity_determining_region_2
 FT region 109..117
 FT /label= complementarity_determining_region_3
 FT region 127 AA.
 PN WO9710354-A1.
 PD 20-MAR-1997.
 PF 11-SEP-1996; J02588.
 PR 11-SEP-1995; JP-232384.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Anazawa H, Furuya A, Hanai N, Iida A, Koike M;
 PI Nakamura K, Takatsu K;
 DR WPI; 97-202249/18.
 DR N-PSDB; T73611.
 PT Antibody against alpha-chain of human interleukin 5 receptor -
 PT useful for diagnosis and treatment of respiratory allergic diseases,
 PT e.g. chronic bronchitis
 PS Example 2; Pages 122-123; 238pp; Japanese.
 CC The present sequence is the light chain variable region of the
 CC murine anti-human interleukin 5 receptor alpha chain (hIL-5R alpha)
 CC monoclonal antibody (MAB) KM1259. KM1259 is produced by the hybridoma
 CC FERM BP-5134, which was prepared by immunising Balb/c mice with
 CC hIL-5R alpha, fusing spleen cells obtained from the mice with mouse
 CC myeloma P3-U1 cells and screening the resultant hybridomas. The MAB
 CC can be used to detect or assay for hIL-5R alpha and cells
 CC expressing it on their surface, especially to diagnose allergic
 CC respiratory diseases, e.g. chronic bronchitis. It can also be used
 CC to treat such diseases.
 SQ Sequence 127 AA;

Query Match 87.2%; Score 34; DB 24; Length 127;
 Best Local Similarity 66.7%; Pred. No. 3.38e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 70 htsrlq 75
 1:1:11
 QY 1 HSSKLQ 6

RESULT 10
 ID W21853 standard; Protein; 127 AA.
 AC W21853;
 DT 19-DEC-1997 (first entry)
 DE Humanised light chain variable region of KM1259 antibody.
 KW Complementarity determining region; CDR; light chain; treatment;
 KW variable region; murine; mouse; human; interleukin 5; IL-5;
 KW receptor; alpha chain; monoclonal antibody; hybridoma; detection;
 KW assay; diagnosis; allergic respiratory disease; humanised;
 KW chronic bronchitis; chimeric.
 OS Chimeric - Mus spp.
 OS Chimeric - Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT peptide 1..20
 FT /label= sig_peptide
 FT peptide 21..127
 FT /label= mat_peptide
 FT region 44..54
 FT /label= complementarity_determining_region_1
 FT region 70..76
 FT /label= complementarity_determining_region_2
 FT region 109..117
 FT /label= complementarity_determining_region_3
 FT region 127 AA.
 PN WO9710354-A1.
 PD 20-MAR-1997.
 PF 11-SEP-1996; J02588.
 PR 11-SEP-1995; JP-232384.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Anazawa H, Furuya A, Hanai N, Iida A, Koike M;
 PI Nakamura K, Takatsu K;
 DR WPI; 97-202249/18.
 DR N-PSDB; T73647.
 PT Antibody against alpha-chain of human interleukin 5 receptor -
 PT useful for diagnosis and treatment of respiratory allergic diseases,
 PT e.g. chronic bronchitis
 PS Example 2; Pages 154-155; 238pp; Japanese.
 CC The present sequence is the humanised light chain variable
 CC region of the murine anti-human interleukin 5 receptor alpha chain
 CC (hIL-5R alpha) monoclonal antibody (MAB) KM1259. KM1259 is produced
 CC by the hybridoma FERM BP-5134, which was prepared by immunising
 CC Balb/c mice with hIL-5R alpha, fusing spleen cells obtained from
 CC the mice with mouse myeloma P3-U1 cells and screening the resultant
 CC hybridomas. The MAB can be used to detect or assay for hIL-5R alpha
 CC and cells expressing it on their surface, especially to diagnose
 CC allergic respiratory diseases, e.g. chronic bronchitis. It can also
 CC be used to treat such diseases.
 SQ Sequence 127 AA;

Query Match 87.2%; Score 34; DB 24; Length 127;
 Best Local Similarity 66.7%; Pred. No. 3.38e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 70 htsrlq 75
 1:1:11
 QY 1 HSSKLQ 6

RESULT 11
 ID W21848 standard; Protein; 127 AA.
 AC W21848;
 DT 19-DEC-1997 (first entry)
 DE Humanised light chain variable region of KM1259 antibody.
 KW Complementarity determining region; CDR; light chain; treatment;
 KW variable region; murine; mouse; human; interleukin 5; IL-5;
 KW receptor; alpha chain; monoclonal antibody; hybridoma; detection;
 KW assay; diagnosis; allergic respiratory disease; humanised;
 KW chronic bronchitis; chimeric.
 OS Chimeric - Mus spp.
 OS Chimeric - Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT peptide 1..20

FT peptide /label= sig_peptide
 FT 21..127
 FT /label= mat_peptide
 FT region 44..54
 FT /label= complementarity_determining_region_1
 FT region 70..76
 FT /label= complementarity_determining_region_2
 FT 109..117
 FT /label= complementarity_determining_region_3
 PN WO9710354-A1.
 PD 20-MAR-1997.
 PF 11-SEP-1996; J02588.
 PR 11-SEP-1995; JP-232384.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Anazawa H, Furuya A, Hanai N, Iida A, Koike M;
 PI Nakamura K, Takatsu K;
 DR WPI: 97-202249/18.
 DR N-PSDB; T73631.
 PT Antibody against alpha-chain of human interleukin 5 receptor -
 PT useful for diagnosis and treatment of respiratory allergic diseases,
 PT e.g. chronic bronchitis
 PS Example 2; Page 141; 238pp; Japanese.
 CC The present sequence is the humanised light chain variable
 CC region of the murine anti-human interleukin 5 receptor alpha chain
 CC (hIL-5R alpha) monoclonal antibody (Mab) KM1259. KM1259 is produced
 CC by the hybridoma FERM BP-5134, which was prepared by immunising
 CC Balb/c mice with hIL-5R alpha, fusing spleen cells obtained from
 CC the mice with mouse myeloma P3-U1 cells and screening the resultant
 CC hybridomas. The Mab can be used to detect or assay for hIL-5R alpha
 CC and cells expressing it on their surface, especially to diagnose
 CC allergic respiratory diseases, e.g. chronic bronchitis. It can also
 CC be used to treat such diseases.
 SQ Sequence 127 AA;

Query Match 87.2%; Score 34; DB 24; Length 127;
 Best Local Similarity 66.7%; Pred. No. 3.38e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 70 htsrlq 75
 Qy 1 HSSKLO 6

RESULT 12
 ID W21852 standard; Protein; 127 AA.
 AC W21852;
 DE Humanised light chain variable region of KM1259 antibody.
 KW Complementarity determining region; CDR; light chain; treatment;
 KW variable region; murine; mouse; human; interleukin 5; IL-5;
 KW receptor; alpha chain; monoclonal antibody; hybridoma; detection;
 KW assay; diagnosis; allergic respiratory disease; humanised;
 KW chronic bronchitis; chimeric.
 OS Chimeric - Mus spp.
 OS Chimeric - Homo sapiens.
 OS Synthetic.
 EH Key Location/Qualifiers
 FT peptide 1..20
 FT /label= sig_peptide
 FT 21..127
 FT /label= mat_peptide
 FT region 44..54
 FT /label= complementarity_determining_region_1
 FT region 70..76
 FT /label= complementarity_determining_region_2
 FT 109..117
 FT /label= complementarity_determining_region_3
 PN WO9710354-A1.
 PD 20-MAR-1997.
 PF 11-SEP-1996; J02588.
 PR 11-SEP-1995; JP-232384.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Anazawa H, Furuya A, Hanai N, Iida A, Koike M;

PI Nakamura K, Takatsu K;
 DR WPI: 97-202249/18.
 DR N-PSDB; T73644.
 PT Antibody against alpha-chain of human interleukin 5 receptor -
 PT useful for diagnosis and treatment of respiratory allergic diseases,
 PT e.g. chronic bronchitis
 PS Example 2; Pages 151-152; 238pp; Japanese.
 CC The present sequence is the humanised light chain variable
 CC region of the murine anti-human interleukin 5 receptor alpha chain
 CC (hIL-5R alpha) monoclonal antibody (Mab) KM1259. KM1259 is produced
 CC by the hybridoma FERM BP-5134, which was prepared by immunising
 CC Balb/c mice with hIL-5R alpha, fusing spleen cells obtained from
 CC the mice with mouse myeloma P3-U1 cells and screening the resultant
 CC hybridomas. The Mab can be used to detect or assay for hIL-5R alpha
 CC and cells expressing it on their surface, especially to diagnose
 CC allergic respiratory diseases, e.g. chronic bronchitis. It can also
 CC be used to treat such diseases.
 SQ Sequence 127 AA;

Query Match 87.2%; Score 34; DB 24; Length 127;
 Best Local Similarity 66.7%; Pred. No. 3.38e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 70 htsrlq 75
 Qy 1 HSSKLO 6

RESULT 13
 ID R10310 standard; Protein; 179 AA.
 AC R10310;
 DE Ovary tissue transcriptional factor DNA clone p2130 product.
 KW Ovary tissue transcriptional factor; DNA construct; probe;
 KW clone p2130.
 OS Lycopersicon esculentum UC82B.
 PN EP-409629-A.
 PD 23-JAN-1991.
 PF 19-JUL-1990; 307926.
 PR 19-JUL-1989; US-382518.
 PA (CALG-) CALGENE INC.
 PI Martineau B, Houck CM;
 DR WPI: 91-024191/04.
 DR N-PSDB; Q10262.
 PT New ovary tissue transcriptional factors - modify transcription
 PT in tomato plant ovaries for use as mol. probes
 PS Disclosure; Fig 1; 21pp; English.
 CC The tomato-derived transcriptional initiation region which regulates
 CC the expression of the sequence corresp. to the p2130 clone is
 CC considered ovary-specific. Sequences hybridisable to the p2130
 CC clone, e.g. probe p27, show abundant mRNA, esp. at the early stages
 CC of anthesis. The message is expressed in ovary integument and ovary
 CC outer pericarp tissue and is not expressed, or at least not readily
 CC detectable, in other tissues or at any other stage of fruit
 CC development. The native function of the amino acid sequence
 CC encoded by the structural gene comprising p2130 is unknown.
 CC See also Q10263-64.
 SQ Sequence 179 AA;

Query Match 87.2%; Score 34; DB 2; Length 179;
 Best Local Similarity 83.3%; Pred. No. 3.38e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 72 hcsklq 77
 Qy 1 HSSKLO 6

RESULT 14
 ID R30779 standard; Protein; 188 AA.
 AC R30779;
 DT 19-MAY-1993 (first entry)
 DE p2130 polypeptide.

KW cDNA; clone p2130; anthesis; tomato; ovary; integument; outer pericarp;
 KW fruit; development; transcription; initiation; region; modulation;
 KW ovary-specific; endogenous; fruit product; exogenous; phenotype.
 OS Lycopersicon esculentum.

FH Key Location/Qualifiers
 FT misc_difference 119 /note= "Nonsense codon"
 FT misc_difference 120 /note= "Nonsense codon"
 FT misc_difference 125 /note= "Nonsense codon"
 FT misc_difference 129 /note= "Nonsense codon"
 FT misc_difference 145 /note= "Nonsense codon"
 FT misc_difference 152 /note= "Nonsense codon"
 FT misc_difference 160 /note= "Nonsense codon"
 FT misc_difference 164 /note= "Nonsense codon"
 FT misc_difference 181 /note= "Nonsense codon"

PN US5175095-A.
 PD 29-DEC-1992.
 PF 19-JUL-1989; 382518.
 PR 19-JUL-1989; US-382518.
 PR 17-JUL-1990; US-554195.
 PA (CALJ) CALGENE INC.
 PI Houck CM, Martineau BM;
 DR WPI: 93-026940/03.
 DR N-PSDB; Q34940.

PT DNA constructs contg. tomato p2130 transcriptional initiation
 PT region - useful for modulation of endogenous fruit prods. and for
 PT prodn. of exogenous prods.

PS Disclosure; Fig 1: 18pp; English.
 CC The sequence represents the polypeptide of cDNA clone p2130. This
 CC sequence is expressed during the early stages of anthesis in tomatoes.
 CC The message is expressed in ovary integument and ovary outer pericarp
 CC tissue. It is not readily detectable in other tissues or at other
 CC stages of fruit development. The transcription initiation region
 CC associated with this gene is therefore considered to be ovary-
 CC specific. The actual function of the p2130 polypeptide is unknown.
 CC The transcription initiation region can be used for modulation of
 CC endogenous fruit products, for production of exogenous products and
 CC for modification of the phenotype of fruit and fruit products.
 SQ Sequence 188 AA;

Query Match 87.2%; Score 34; DB 6; Length 188;
 Best Local Similarity 83.3%; Pred. No. 3.38e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 72 hcsklq 77
 | | | | |
 QY 1 HSSKLQ 6

RESULT 15

ID W50891 standard; Protein: 3084 AA.
 AC W50891;
 DT 07-DEC-1998 (first entry)
 DE Mouse laminin A chain.

KW Laminin; mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
 KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
 KW malignancy; Familial Mediterranean Fever; multiple myeloma;
 KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
 KW Gertsman-Straussler syndrome; kuru; scrapie; haemodialysis;
 KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
 KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;

KW therapy.
 KW Mus sp.
 FH Key Location/Qualifiers
 FT Domain 2746..2922

FT Region /note= "fourth globular domain repeat (Claim 13)"
 FT 2690..2700
 FT /note= "beta-amyloid protein binding region
 FT (Claim 12)"
 PN W09815179-A1.
 PD 16-APR-1998.
 PF 08-OCT-1997; U18145.
 PR 08-OCT-1996; US-027981.
 PA (UNIW) UNIV WASHINGTON.
 PI Castillo G, Snow AD;
 DR WPI: 98-240534/21.

PT Use of laminin and fragments - for developing products for use in
 PT the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
 PT disease or CJD
 PS Claim 15; Page 74-79; 132pp; English.
 CC This is the amino acid sequence of the mouse laminin A chain. The
 CC primary object of the invention is to use laminin, laminin-derived
 CC protein fragments and/or laminin-derived polypeptides as potent
 CC inhibitors of amyloid formation, deposition, accumulation and/or
 CC persistence in Alzheimer's disease and other amyloidoses. The
 CC laminin products (see W50888-98) may include mouse or human laminin
 CC A or Al chain, laminin B1 or B2 chain, laminin A2 chain (merosin),
 CC laminin G1 chain, the globular repeats of the laminin A1 chain and
 CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC of laminin or its fourth globular repeat in vivo provides a method
 CC for in vivo inhibition of beta-amyloid amyloidosis. The products
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidoses such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (AA
 CC amyloid or inflammation-association amyloidosis), the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gertsmann-Straussler
 CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and Familial Amyloidotic Polynuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).
 SQ Sequence 3084 AA;

Query Match 87.2%; Score 34; DB 35; Length 3084;
 Best Local Similarity 83.3%; Pred. No. 3.38e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1764 hseklq 1769
 | | | | |
 QY 1 HSSKLQ 6

Search completed: Thu Oct 28 11:19:48 1999
 Job time : 21 secs.

WIRE (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Oct 28 11:20:06 1999; MasPar time 2.92 Seconds
Tabular output not generated. 82.440 Million cell updates/sec

Title: >US-09-081-707-7
Description: (1-6) from US09081707.ppe
Perfect Score: 39
Sequence: 1 HSSKIQ 6

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 17.857; Variance 14.939; scale 1.195

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	ID	Description	Pred. No.
1	39	100.0	147	2 S13157 hemoglobin - polychaete	1.54e+00
2	39	100.0	582	2 A43412 semenogelin II precursor	1.54e+00
3	37	94.9	91	2 E65011 hypothetical protein	5.99e+00
4	37	94.9	278	2 H65069 hypothetical protein	5.99e+00
5	37	94.9	883	2 A71434 probable RNA helicase	5.99e+00
6	36	92.3	147	2 B36529 hemoglobin P3 - polyc	1.16e+01
7	36	92.3	509	2 T01344 hypothetical protein	1.16e+01
8	36	92.3	531	2 B54096 flavin-containing mon	1.16e+01
9	36	92.3	533	1 S18380 dimethylalliline monoo	1.16e+01
10	36	92.3	675	3 T00013 DAP-1 beta protein	1.16e+01
11	36	92.3	692	2 T00025 PSD-95 binding protei	1.16e+01
12	36	92.3	744	2 S45060 outer capsid spike pr	1.16e+01
13	36	92.3	744	2 S45061 outer capsid spike pr	1.16e+01
14	36	92.3	977	3 T00014 DAP-1 alpha protein	1.16e+01
15	35	89.7	2067	2 A42854 probable spindle pole	2.21e+01
16	34	87.2	105	2 S57809 gamma-thionin-like pr	4.14e+01
17	34	87.2	147	2 A36529 hemoglobin P2 - polyc	4.14e+01
18	34	87.2	265	2 S63532 NAD(P)H-quinone oxido	4.14e+01
19	34	87.2	294	2 S58472 surface antigen - Ent	4.14e+01
20	34	87.2	579	2 T02574 hypothetical protein	4.14e+01
21	34	87.2	585	2 G64220 ATP-binding protein m	4.14e+01
22	34	87.2	1377	2 A38926 DNA-binding protein c	4.14e+01
23	34	87.2	3084	1 MMWSA_laminin alpha-1 chain	4.14e+01

RESULT 1
ENTRY S13157 #type complete
TITLE hemoglobin - polychaete (Glycera dibranchiata)
ORGANISM #formal_name Glycera dibranchiata #common_name bloodworm
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Feb-1998

ACCESSIONS S13157
REFERENCE S13157
#authors Zafar, R.S.; Chow, L.H.; Stern, M.S.; Vinogradov, S.N.; Walz, D.A.

#journal Biochim. Biophys. Acta (1990) 1041:117-122
#title The heterogeneity of the polymeric intracellular hemoglobin of Glycera dibranchiata and the cDNA-derived amino acid sequence of one component.

#cross-references MUID:91091412
#accession S13157
##status preliminary
##molecule_type mRNA
##residues 1-147 ##label ZAF

CLASSIFICATION #superfamily globin; globin homology
SUMMARY #length 147 #molecular-weight 16029 #checksum 7221

Query Match 100.0%; Score 39; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.54e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 57 HSSKIQ 62
|||||
Qy 1 HSSKIQ 6

RESULT 2
ENTRY A43412 #type complete
TITLE semenogelin II precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-May-1998

ACCESSIONS A43412; B31489; A45295; S29156; S68765; S68762
REFERENCE A43412
#authors Ulvback, M.; Lazure, C.; Lilja, H.; Spurr, N.K.; Rao, V.V.; Loeffler, C.; Hansmann, I.; Lundwall, A.
#journal J. Biol. Chem. (1992) 267:18080-18084
#title Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20.
#cross-references MUID:92388176
#accession A43412

33 84.6 160 2 C35542 ribosomal protein S11 7.65e+01
33 84.6 179 2 S74360 hypothetical protein 7.65e+01
33 84.6 210 2 A49213 major surface protein 7.65e+01
33 84.6 211 2 C71482 probable endonuclease 7.65e+01
33 84.6 252 2 P04259 ferritin associated p 7.65e+01
33 84.6 273 2 S59305 hypothetical protein 7.65e+01
33 84.6 276 2 D64014 hypothetical protein 7.65e+01
33 84.6 276 2 E71705 hypothetical protein 7.65e+01
33 84.6 369 2 S16482 docking protein - Sul 7.65e+01
33 84.6 369 2 S53703 dpa protein - Sulfolo 7.65e+01
33 84.6 366 2 A40679 transcription enhance 7.65e+01
33 84.6 585 2 E64892 probable membrane pro 7.65e+01
33 84.6 589 2 I38598 zinc finger protein 2 7.65e+01
33 84.6 617 2 T02121 hypothetical protein 7.65e+01
33 84.6 631 2 S36505 El protein - human pa 7.65e+01
33 84.6 664 2 S58162 hypothetical protein 7.65e+01
33 84.6 739 2 A46159 interferon-dependent 7.65e+01
33 84.6 779 2 T01304 hypothetical protein 7.65e+01
33 84.6 839 1 GNNYS2 genome polypeptide - 7.65e+01
33 84.6 896 1 GNLJGH pol polypeptide - hum 7.65e+01
33 84.6 896 1 GNLJCN pol polypeptide - hum 7.65e+01
33 84.6 1173 2 I50620 prockr2 - chicken (fr 7.65e+01

ALIGNMENTS

```

##molecule_type DNA
##residues 1-582 #label ULV
##cross-references GB:M81631; NID:g307417; PID:g307418
##note sequence extracted from NCBI backbone (NCBI:112887, NCBIP:112889)

REFERENCE
#authors A31489
#journal Lilja, H.; Abrahamson, P.A.; Lundwall, A.
#title J. Biol. Chem. (1989) 264:1894-1900
#title Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa.
#cross-references MUID:89109215
#accession B31489
##status nucleic acid sequence not shown
##molecule_type mRNA
##residues 214-278, 'Y', 280-281 #label LIL

REFERENCE
#authors A45295
#journal Lilja, H.; Lundwall, A.
#title Proc. Natl. Acad. Sci. U.S.A. (1992) 89:4559-4563
#title Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein.
#cross-references MUID:92262479
#accession A45295
##molecule_type mRNA
##residues 3-582 #label LI2
##cross-references GB:M81652

REFERENCE
#authors S29155
#journal Schneider, K.; Kausler, W.; Tripiet, D.; Jouvenal, K.; Spiteiller, G.
#title Biol. Chem. Hoppe-Seyler (1989) 370:353-356
#title Isolation and structure determination of two peptides occurring in human seminal plasma.
#accession S29156
##molecule_type protein
##residues 390, 'E', 392-396, 'E', 398-400, 'EW', 403-405, 'C', 407, 'DE', #label SCH
##note this report is of a secondary sequence determined simultaneously with the sequence with accession number S29155 (see entry WTHUB); as a secondary sequence, it should be considered less than fully reliable

REFERENCE
#authors S68765
#journal Kise, H.; Nishioka, J.; Kawamura, J.; Suzuki, K.
#title Eur. J. Biochem. (1996) 238:88-96
#title Characterization of semenogelin II and its molecular interaction with prostate-specific antigen and protein C inhibitor.
#cross-references MUID:96248425
#accession S68765
##molecule_type protein
##residues 105-107, 'K', 109-111, 113-122, 260-269, 280-283 #label KIS

REFERENCE
#authors S68761
#journal Malm, J.; Hellman, J.; Magnusson, H.; Laurell, C.B.; Lilja, H.
#title Eur. J. Biochem. (1996) 238:48-53
#title Isolation and characterization of the major gel proteins in human semen, semenogelin I and semenogelin II.
#accession S68762
##molecule_type protein
##residues 420-421, 'G', 423-423 #label MAL

GENETICS
#gene GDB:SEMG2
##cross-references GDB:132657; OMIM:182141
#map_position 20q12-20q13.1
#introns 26/1
CLASSIFICATION
#superfamily semenogelin
KEYWORDS
duplication; glycoprotein; semen; seminal vesicle; tandem repeat
FEATURE
#domain signal sequence #status predicted #label SIG\
#product semenogelin II #status predicted #label MAT\
#binding_site carbohydrate (Asn) (covalent) #status predicted
#length 582 #molecular-weight 65444 #checksum 9704
SUMMARY

```

```

Query Match 100.0%; Score 39; DB 2; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.54e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 230 HSKLQ 235
QY 1 HSKLQ 6

RESULT 3
ENTRY #type complete
TITLE hypothetical protein b2376 - Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 14-Nov-1997
ACCESSION E65011
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession E65011
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-91 #label BLAT
##cross-references GB:AE000326; GB:U00096; NID:g1788718; PID:g2367134; UWGP:b2376
##experimental_source strain K-12, substrain MG1655
SUMMARY #length 91 #molecular-weight 10162 #checksum 3897

Query Match 94.9%; Score 37; DB 2; Length 91;
Best Local Similarity 83.3%; Pred. No. 5.99e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 48 HSKLQ 53
QY 1 HSKLQ 6

RESULT 4
ENTRY #type complete
TITLE hypothetical protein b2863 - Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 14-Nov-1997
ACCESSION H65069
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession H65069
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-278 #label BLAT
##cross-references GB:AE000369; GB:U00096; NID:g2367168; PID:g1789227; UWGP:b2863
##experimental_source strain K-12, substrain MG1655
SUMMARY #length 278 #molecular-weight 31319 #checksum 4142

Query Match 94.9%; Score 37; DB 2; Length 278;
Best Local Similarity 83.3%; Pred. No. 5.99e+00;

```



```

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 77 HSAKIQ 82
   ||:||||
QY 1 HSSKIQ 6

RESULT 5
ENTRY #type complete
TITLE probable RNA helicase - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress

#variety
DATE 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
A71434
A71434
A71400
Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.;
Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.;
Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
Leclercq, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.;
Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger,
M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
A.; Vouklatou, E.; Milioni, D.; Hatzopoulos, P.;
Paravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
Reichman, S.; Ansoorge, W.; Cooke, R.; Berger, C.; Delseny,
M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
Schueller, C.; Chalwatzis, N.
#journal Nature (1998) 391:485-488
#title Analysis of 1.9 Mb of contiguous sequence from chromosome 4
of Arabidopsis thaliana.
#cross-references MUID:98121113
#accession A71434
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-883 #label BEV
#cross-references GB:297341; NID:g2244991; PID:e327007; PID:g2245023
GENETICS
#map_position 4COP9-4G3845
#length 883 #molecular-weight 100986 #checksum 5437
SUMMARY

Query Match 94.9%; Score 37; DB 2; Length 883;
Best Local Similarity 83.3%; Pred. No. 5.99e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 771 HSAKIQ 776
   ||:||||
QY 1 HSSKIQ 6

RESULT 6
ENTRY #type complete
TITLE hemoglobin P3 - polychaete (Glycera dibranchiata)
ORGANISM #formal_name Glycera dibranchiata #common_name bloodworm
DATE 12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change
16-Dec-1998
A36529
A36529
Zafar, R.S.; Chow, L.H.; Stern, M.S.; Scully, J.S.; Sharma,
P.R.; Vinogradov, S.N.; Walz, D.A.
#journal J. Biol. Chem. (1990) 265:21843-21851
#title The cDNA sequences encoding two components of the polymeric
fraction of the intracellular hemoglobin of Glycera
dibranchiata.
#cross-references MUID:91072390
#accession B36529
#status preliminary

#molecule_type mRNA
#residues 1-147 #label ZAF
#cross-references GB:M55444; GB:J05691; NID:g159146; PID:g159147
CLASSIFICATION #superfamily globin; globin homology
FEATURE 3-142
SUMMARY #domain globin homology #label GLB
#length 147 #molecular-weight 15977 #checksum 8871

Query Match 92.3%; Score 36; DB 2; Length 147;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 57 HNSKIQ 62
   ||:||||
QY 1 HSSKIQ 6

RESULT 7
ENTRY #type complete
TITLE hypothetical protein F6N15.5 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
T01344
T01344
Z14297
Ryan, E.; Edwards, J.; Pape, K.
#submitters submitted to the EMBL Data Library, May 1998
#description The sequence of A. thaliana F6N15.
#accession T01344
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-509 #label RYA
#cross-references EMBL:AF069299; NID:g3193311; PID:g3193321
GENETICS
#map_position IV
#introns 50/2; 342/3; 399/3; 496/2
#note F6N15.5
SUMMARY #length 509 #molecular-weight 58562 #checksum 4630

Query Match 92.3%; Score 36; DB 2; Length 509;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 376 HSPKIQ 381
   ||:||||
QY 1 HSSKIQ 6

RESULT 8
ENTRY #type complete
TITLE flavin-containing monooxygenase (EC 1.14.-.-) FMO3 - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
08-Sep-1997
B54096
B54096
Burnett, V.L.; Lawton, M.P.; Philpot, R.M.
#authors J. Biol. Chem. (1994) 269:14314-14322
#journal Cloning and sequencing of flavin-containing monooxygenases
FMO3 and FMO4 from rabbit and characterization of FMO3.
#cross-references MUID:94245761
#accession B54096
#status preliminary
#molecule_type mRNA
#residues 1-531 #label BUR
#cross-references GB:LI0391; NID:g349533; PID:g349534
CLASSIFICATION #superfamily dimethylalanine monooxygenase (N-oxide-forming)
oxidoreductase
KEYWORDS #length 531 #molecular-weight 59815 #checksum 8965
SUMMARY

Query Match 92.3%; Score 36; DB 2; Length 531;

```

```
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 83 HNSKLG 88
   1 HSSKLG 6

RESULT 9
ENTRY   #type complete
TITLE   dimethylalanine monooxygenase (N-oxide-forming) (EC
        1.14.13.8), hepatic 2 - rabbit
ALTERNATE_NAMES
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
        rabbit
DATE     30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
        13-Mar-1997
ACCESSIONS
REFERENCE S18380; B32914
#authors Ozols, J.
#journal Arch. Biochem. Biophys. (1991) 290:103-115
#title   Multiple forms of liver microsomal flavin-containing
        monooxygenases: complete covalent structure of form 2.
#cross-references MUID:91378577
#accession S18380
#molecule_type protein
#residues 1-533 #label OZ01
REFERENCE A32914
#authors Ozols, J.
#journal Biochem. Biophys. Res. Commun. (1989) 163:49-55
#title   Liver microsomes contain two distinct NADPH-monoxygenases
        with NH-2-terminal segments homologous to the flavin
        containing NADPH monooxygenase of Pseudomonas fluorescens.
#cross-references MUID:89374273
#accession B32914
#molecule_type protein
#residues 1-32 #label OZO2
COMMENT  This enzyme is involved in the metabolism of many drugs,
        pesticides, and other foreign compounds, including xenobiotics,
        by catalyzing the NADPH-dependent oxidation of various
        substrates.
CLASSIFICATION #superfamily dimethylalanine monooxygenase (N-oxide-forming)
KEYWORDS      FAD; flavoprotein; microsome; monooxygenase; NADP;
        oxidoreductase
FEATURE
3-31
185-213
SUMMARY      #region beta-alpha-beta FAD nucleotide-binding fold\
        #region beta-alpha-beta NADP nucleotide-binding fold
        #length 533 #molecular-weight 60096 #checksum 1027

Query Match 92.3%; Score 36; DB 1; Length 533;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 82 HNSKLG 87
   1 HSSKLG 6

RESULT 10
ENTRY   #type complete
TITLE   DAP-1 beta protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     19-Jan-1999 #sequence_revision 19-Jan-1999 #text_change
        19-Jan-1999
ACCESSIONS
REFERENCE T00013
#authors Satoh, K.; Yanai, H.; Senda, T.; Kohu, K.; Nakamura, T.;
        Okumura, N.; Matsumine, A.; Kobayashi, S.; Toyoshima, K.;
        Akiyama, T.
#journal Genes to Cells (1997) 2:415-424
#title   DAP-1, a novel protein that interacts with the guanylate
        kinase-like domains of hDLG and PSD-95.
#accession T00013
```

```
##status preliminary
##residues 1-675 #label SAT
##cross-references EMBL:AB000276; NID:d1173041; PID:d1024134
SUMMARY #length 675 #molecular-weight 75260 #checksum 1624

Query Match 92.3%; Score 36; DB 3; Length 675;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 94 HSPKLG 99
   1 HSSKLG 6

RESULT 11
ENTRY   #type complete
TITLE   PSD-95 binding protein - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE     22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
        22-Jan-1999
ACCESSIONS
REFERENCE T00025
#authors Kawashima, N.; Takamiya, K.; Sun, J.; Kitabatake, A.; Sobue,
        K.
#journal FEBS Lett. (1997) 418:301-304
#title   Differential expression of isoforms of PSD-95 binding protein
        (GRAP/SAPAP1) during rat brain development.
#accession T00025
##status preliminary; translated from GB/EMBL/DBBJ
##molecule_type mRNA
#residues 1-692 #label KAW
##cross-references EMBL:AB003594; NID:d1177606; PID:d1025176
SUMMARY #length 692 #molecular-weight 76991 #checksum 9411

Query Match 92.3%; Score 36; DB 2; Length 692;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 102 HSPKLG 107
   1 HSSKLG 6

RESULT 12
ENTRY   #type complete
TITLE   outer capsid spike protein - human rotavirus C
ORGANISM #formal_name human rotavirus C
DATE     13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
        08-Sep-1997
ACCESSIONS
REFERENCE S45060
#authors Fielding, P.A.; Lambden, P.R.; Caul, E.O.; Clarke, I.N.
#submission submitted to the EMBL Data Library, May 1994
#description Molecular cloning, complete nucleotide sequence and coding
        assignment of the outer capsid spike protein (VP4) from a
        human group C rotavirus: comparison with th VP4 gene of a
        Brazilian isolate.
#accession S45060
##status preliminary
##molecule_type DNA
#residues 1-744 #label FIE
##cross-references EMBL:X79441; NID:g495228; PID:g495229
CLASSIFICATION #superfamily rotavirus outer layer protein VP3
SUMMARY #length 744 #molecular-weight 84036 #checksum 6403

Query Match 92.3%; Score 36; DB 2; Length 744;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 410 HSSRLQ 415
   1 HSSKLG 6
```

```
RESULT 13
ENTRY   #type complete
TITLE   outer capsid spike protein - human rotavirus C
ORGANISM #formal_name human rotavirus C
DATE    13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
        08-Sep-1997
ACCESSIONS
REFERENCE S45061
          S45060
          Fielding, P.A.; Lambden, P.R.; Caul, E.O.; Clarke, I.N.
          submitted to the EMBL Data Library, May 1994
          Molecular cloning, complete nucleotide sequence and coding
          assignment of the outer capsid spike protein (VP4) from a
          human group C rotavirus: comparison with th VP4 gene of a
          brazilian isolate.
          #accession S45061
          ##status preliminary
          ##molecule_type DNA
          ##residues 1-744 #label FIE
          ##cross-references EMBL:X79442; NID:g495230; PID:g495231
CLASSIFICATION #superfamily rotavirus outer layer protein VP3
SUMMARY        #length 744 #molecular-weight 84080 #checksum 8279

Query Match      92.3%; Score 36; DB 2; Length 744;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 410 HSRRLQ 415
   |||:|
QY 1 HSKLQ 6

RESULT 14
ENTRY   #type complete
TITLE   DAP-1 alpha protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE    19-Jan-1999 #sequence_revision 19-Jan-1999 #text_change
        19-Jan-1999
ACCESSIONS
REFERENCE T00014
          Z14044
          Satoh, K.; Yanai, H.; Senda, T.; Kohu, K.; Nakamura, T.;
          Okumura, N.; Matsumine, A.; Kobayashi, S.; Toyoshima, K.;
          Akiyama, T.
          Genes to Cells (1997) 2:415-424
          DAP-1, a novel protein that interacts with the guanylate
          kinase-like domains of hDLG and PSD-95.
          #accession T00014
          ##status preliminary
          ##residues 1-977 #label SAT
          ##cross-references EMBL:A800027; NID:d1173042; PID:d1024135
SUMMARY        #length 977 #molecular-weight 108872 #checksum 4168

Query Match      92.3%; Score 36; DB 3; Length 977;
Best Local Similarity 83.3%; Pred. No. 1.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 396 HSPKLQ 401
   |||:|
QY 1 HSKLQ 6

RESULT 15
ENTRY   #type complete
TITLE   probable spindle pole body component bimB - Emericella
          nidulans
ALTERNATE_NAMES cell division-associated protein bimB
ORGANISM #formal_name Emericella nidulans, Aspergillus nidulans
DATE    10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
        10-Oct-1997
ACCESSIONS
REFERENCE A42854
          S27411
          May, G.S.; McGoldrick, C.A.; Holt, C.L.; Denison, S.H.
          J. Biol. Chem. (1992) 267:15737-15743

#title The bimB3 mutation of Aspergillus nidulans uncouples DNA
#cross-references MUID:92348436
#accession A42854
##molecule_type mRNA
##residues 1-2067 #label MAY
##note sequence extracted from NCBI backbone (NCBIN:112444,
        NCBI:P:112445)
REFERENCE S27411
#authors May, G.S.; McGoldrick, C.A.; Denison, S.H.; Holt, C.L.
#submission submitted to the EMBL Data Library, February 1992
#description The bimB3 mutation of Aspergillus nidulans uncouples DNA
        replication from the completion of mitosis.
#accession S27411
##status preliminary
##molecule_type DNA
##residues 1-2010, 'PLSIG', 2016 #label MA2
##cross-references EMBL:M83232; NID:g168020; PID:g168021
##note this sequence differs by a frameshift after residue 2010
KEYWORDS nucleus
SUMMARY        #length 2067 #molecular-weight 227924 #checksum 9014

Query Match      89.7%; Score 35; DB 2; Length 2067;
Best Local Similarity 66.7%; Pred. No. 2.21e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1665 HAAKLQ 1670
   |||:|
QY 1 HSKLQ 6

Search completed: Thu Oct 28 11:20:20 1999
Job time : 14 secs.
```

THIS PAGE BLANK (USPTO)

W A R E (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Oct 28 11:20:37 1999; MasPar time 2.03 Seconds
Tabular output not generated. 83.380 Million cell updates/sec

Title: >US-09-081-707-7
Description: (1-6) from US09081707.pep
Perfect Score: 39
Sequence: 1 HSSKIQ 6

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 18.322; Variance 13.214; scale 1.387

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	39	100.0	147	1 GLP1_GLYDI	GLOBIN, MAJOR POLYMERIC	3.93e-01
2	39	100.0	380	1 PEX2_YARLI	PEROXISOME ASSEMBLY PR	3.93e-01
3	39	100.0	582	1 SEM2_HUMAN	SEMNOCHELIN II PRECURS	3.93e-01
4	39	100.0	706	1 SEM2_MACMU	SEMNOCHELIN II PRECURS	3.93e-01
5	37	94.9	91	1 YPDI_ECOLI	HYPOTHETICAL 10.2 KD L	1.81e+00
6	36	92.3	147	1 GLP3_GLYDI	GLOBIN, POLYMERIC COMP	3.80e+00
7	36	92.3	530	1 FM03_RABIT	DIMETHYLANILINE MONOX	3.80e+00
8	36	92.3	954	1 DRP2_HUMAN	DYSTROPHIN-RELATED PRO	3.80e+00
9	35	89.7	2067	1 BMB_EHANI	CELL DIVISION-ASSOCIAT	7.83e+00
10	34	87.2	147	1 GLP2_GLYDI	GLOBIN, POLYMERIC COMP	1.58e+01
11	34	87.2	415	1 TRMU_SCHPO	PROBABLE TRNA (5-METHY	1.58e+01
12	34	87.2	585	1 Y187_MYCGE	HYPOTHETICAL ABC TRANS	1.58e+01
13	34	87.2	1377	1 CID_DROME	CUBITUS INTERRUPTUS DO	1.58e+01
14	34	87.2	1524	1 Y133_HUMAN	HYPOTHETICAL PROTEIN K	1.58e+01
15	34	87.2	2245	1 MYSJ_DICDI	MYOSIN IJ HEAVY CHAIN	1.58e+01
16	34	87.2	3084	1 LMA1_MOUSE	LAMININ ALPHA-1 CHAIN	1.58e+01
17	33	84.6	79	1 KRUI_DROME	KRUEPPEL HOMOLOGOUS PR	3.15e+01
18	33	84.6	140	1 HBB_RANES	HEMOGLOBIN BETA CHAIN	3.15e+01
19	33	84.6	159	1 R11B_ARATH	40S RIBOSOMAL PROTEIN	3.15e+01
20	33	84.6	160	1 R11A_ARATH	40S RIBOSOMAL PROTEIN	3.15e+01
21	33	84.6	179	1 Y053_SYNY3	HYPOTHETICAL 20.1 KD P	3.15e+01
22	33	84.6	255	1 EI28_DROME	ECDYSONE-INDUCED PROTE	3.15e+01
23	33	84.6	273	1 YM92_YEAST	HYPOTHETICAL 30.6 KD P	3.15e+01

24	33	84.6	276	1 Y842_HAEIN	HYPOTHETICAL PROTEIN H	3.15e+01
25	33	84.6	369	1 DOCK_SULAC	PROBABLE SIGNAL RECOGN	3.15e+01
26	33	84.6	451	1 CEB_DROVI	CCAAT/ENHANCER BINDING	3.15e+01
27	33	84.6	493	1 UDPE_NPVCD	ECDYSTEROID UDP-GLUCOS	3.15e+01
28	33	84.6	523	1 GLI2_HUMAN	ZINC FINGER PROTEIN GL	3.15e+01
29	33	84.6	534	1 FMO3_MOUSE	DIMETHYLANILINE MONOX	3.15e+01
30	33	84.6	589	1 Z132_HUMAN	ZINC FINGER PROTEIN 13	3.15e+01
31	33	84.6	631	1 VEL_HPV30	REPLICATION PROTEIN E1	3.15e+01
32	33	84.6	749	1 STAL_MOUSE	SIGNAL TRANSDUCER AND	3.15e+01
33	33	84.6	750	1 STAL_HUMAN	SIGNAL TRANSDUCER AND	3.15e+01
34	33	84.6	777	1 YAZ1_SCHPO	HYPOTHETICAL 88.2 KD P	3.15e+01
35	33	84.6	839	1 POLG_HPAVT	GENOME POLYPROTEIN [CO	3.15e+01
36	33	84.6	896	1 POL_HTLIA	POL POLYPROTEIN [CONTA	3.15e+01
37	33	84.6	896	1 POL_HTLIC	POL POLYPROTEIN [CONTA	3.15e+01
38	33	84.6	1195	1 CHS4_NEUCR	CHITIN SYNTHASE 4 (EC	3.15e+01
39	33	84.6	1381	1 YBE7_YEAST	HYPOTHETICAL 150.8 KD	3.15e+01
40	33	84.6	1385	1 RRPO_PIAMV	RNA REPLICATION PROTEI	3.15e+01
41	33	84.6	1485	1 Z165_HUMAN	ZINC FINGER PROTEIN 16	6.13e+01
42	32	82.1	489	1 UDPE_NPVOP	ECDYSTEROID UDP-GLUCOS	6.13e+01
43	32	82.1	635	1 YGC3_YEAST	HYPOTHETICAL 70.6 KD P	6.13e+01
44	32	82.1	732	1 YMW1_CABEL	HYPOTHETICAL 81.8 KD P	6.13e+01
45	32	82.1	1181	1 Y102_METJA	HYPOTHETICAL PROTEIN M	6.13e+01

ALIGNMENTS

RESULT 1 STANDARD; PRT; 147 AA.
ID GLP1_GLYDI
AC P23216;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE GLOBIN, MAJOR POLYMERIC COMPONENT PL.
OC GLYCERA DIBRANCHIATA (BLOODWORM).
OC EUKARYOTA; METAZOA; ANNELIDA; POLYCHAETA; PHYLLODOCIDA; GLYCERIDAE;
OC GLYCERA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-27; 98-102 AND 119-145.
RX MEDLINE; 91091412.
RA ZAFAR R.S., CHOW L.H., STERN M.S., VINOGRADOV S.N., WALZ D.A.;
RT "The heterogeneity of the polymeric intracellular hemoglobin of
RT Glycera dibranchiata and the cDNA-derived amino acid sequence of one
RT Component."
RL BIOCHIM. BIOPHYS. ACTA 1041:117-122(1990).
CC -!- SUBUNIT: POLYMER.
CC -!- THIS PROTEIN IS ONE OF AT LEAST SIX COMPONENTS IN THE POLYMERIC
CC FRACTION OF GLYCERA HEMOGLOBIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54060; G9347; -
DR PIR; S13157; S13157.
DR PROSITE; PS01033; GLOBIN; 1.
DR PFAM; PF00042; globin; 1.
DR HSP; P02216; 1HGB.
KW HEME; OXYGEN TRANSPORT; RESPIRATORY PROTEIN.
SQ SEQUENCE 147 AA; 16019 MW; 61C9888B CRC32;

Query Match 100.0%; Score 39; DB 1; Length 147;
Best Local Similarity 100.0%; Pred. No. 3.93e-01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 57 HSSKIQ 62
Oy 1 HSSKIQ 6

RESULT 2
ID PEX2_YARLI STANDARD; PRT; 380 AA.
AC Q99155;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PEROXISOME ASSEMBLY PROTEIN PAYS (PEROXIN-2).
GN PEX2 OR PAYS.
OS YAROWIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYTES; SACCHAROMYCETES;
OC DIPODASCACEAE; YAROWIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96355353.
RA EITZEN G.A., TITORENKO V.I., SMITH J.J., VEENHUIS M.,
RA SZILARD R.K., RACHUBINSKI R.A.;
RT "The Yarrowia lipolytica gene PAYS encodes a peroxisomal integral
RT membrane protein homologous to the mammalian peroxisome assembly
RT factor PAF-1.";
RL J. BIOL. CHEM. 271:20300-20306(1996).
CC -!- FUNCTION: REQUIRED FOR THE IMPORT OF SEVERAL PROTEINS INTO
CC PEROXISOMES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL
CC (POTENTIAL).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U43081; G1209744; -
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR PFAM; PF00097; zf-C3HC4; 1.
KW TRANSMEMBRANE; PEROXISOME; ZINC-FINGER.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
FT ZN_FING 265 322 C3HC4-TYPE.
FT DOMAIN 277 280 POLY-GLY.
FT DOMAIN 354 380 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 360 385 POLY-GLU.
FT DOMAIN 367 372 POLY-ASP.
SQ SEQUENCE 380 AA; 42727 MW; 489F7120 CRC32;
Query Match 100.0%; Score 39; DB 1; Length 380;
Best Local Similarity 100.0%; Pred. No. 3.93e-01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 92 HSSKLQ 97
QY 1 HSSKLQ 6
RESULT 3
ID SEM2_HUMAN STANDARD; PRT; 582 AA.
AC Q02383;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SEMENOGELIN II PRECURSOR (SGII).
GN SEMG2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEMINAL VESICLE;
RX MEDLINE; 92262479.
RA LUNDWALL A., LILJA H.;
RT "Molecular cloning of epididymal and seminal vesicular transcripts

RT encoding a semenogelin-related protein.";
RL PROC. NATL. ACAD. SCI. U.S.A. 89:4559-4563(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92388176.
RA ULVSBACK M., LAZURE C., LILJA H., SPURR N.K., RAO V.V., LOEFFLER C.,
RA HANSMANN I., LUNDWALL A.;
RT "Gene structure of semenogelin I and II. The predominant proteins in
RT human semen are encoded by two homologous genes on chromosome 20.";
RL J. BIOL. CHEM. 267:18080-18084(1992).
CC -!- FUNCTION: THREE SEMENOGELIN PROTEINS ARE FOUND IN HUMAN SEMEN,
CC OF WHICH THE 52 KD SEMENOGELIN I IS THE MOST ABUNDANT. THEY
CC PARTICIPATE IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE
CC ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS
CC OF SEMENOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY
CC CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS
CC THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE
CC PROTEASE.
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEXES OF SEMENOGELIN I AND THE OTHER
CC SEMENOGELIN PROTEINS, THE 71- AND 76-KD POLYPEPTIDES.
CC -!- TISSUE SPECIFICITY: SEMINAL VESICLES, AND TO A MUCH LESSER
CC EXTENT, EPIDIDYMS.
CC -!- PTM: SEMENOGELIN II IS THOUGHT TO FORM BOTH THE 71-KD POLYPEPTIDE
CC AND, IN ITS GLYCOSYLATED FORM, THE 76-KD POLYPEPTIDE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81651; G307418; -
DR EMBL; M81652; G338239; -
DR EMBL; 247556; E82662; -
DR PIR; A43412; A43412.
DR MIM; 182141; -
KW SEMEN; SEMINAL VESICLE; REPEAT; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 23
FT CHAIN 24 582 SEMENOGELIN II.
FT DOMAIN 70 559 REPEAT-RICH REGION.
FT REPEAT 70 129 3-1.
FT REPEAT 141 200 2-1.
FT REPEAT 201 260 2-2.
FT DOMAIN 261 500 4 X 60 AA TANDEM REPEATS, TYPE I.
FT REPEAT 501 559 3-2.
FT CARBOHYD 272 272 PROBABLE.
SQ SEQUENCE 582 AA; 65445 MW; DD20304E CRC32;
Query Match 100.0%; Score 39; DB 1; Length 582;
Best Local Similarity 100.0%; Pred. No. 3.93e-01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 230 HSSKLQ 235
QY 1 HSSKLQ 6
RESULT 4
ID SEM2_MACMU STANDARD; PRT; 706 AA.
AC Q95196;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SEMENOGELIN II PRECURSOR (SGII).
GN SEMG2.
OS MACACA MULATTA (RHESUS MACAQUE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97274635.

RA ULVSBACK M., LUNDWALL A.:
RT "Cloning of the semenogelin II gene of the rhesus monkey.
RT Clonings of 360 bp extend the coding region in man, rhesus monkey
RL and baboon."
RL EUR. J. BIOCHEM. 245:25-31(1997).
CC -1- FUNCTION: PARTICIPATE IN THE FORMATION OF A GEL MATRIX ENTRAPPING
CC THE ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X92589; E03870;
KW SEMEN; SEMINAL VESICLE; REPEAT; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 706 SEMENOGELIN II.
SQ SEQUENCE 706 AA; 79875 MW; 9D5897A8 CRC32;

Query Match 100.0%; Score 39; DB 1; Length 706;
Best Local Similarity 100.0%; Pred. No. 3.93e-01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 234 HSKLQ 239
QY 1 HSKLQ 6

RESULT 5
ID YPDI_ECOLI STANDARD; PRT; 91 AA.
AC O32528; (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 10.2 KD LIPOPROTEIN IN EVGS-DBG REGION PRECURSOR.
GN YPDI
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
CC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., ROSE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.
RT "The complete genome sequence of Escherichia coli K-12."
RL SCIENCE 277:1453-1474(1997).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000326; G2367134;
DR ECOGENE; EG14376; YPDI.
DR PROSITE; PS00013; PROKAR.LIPOPROTEIN: 1.
KW HYPOTHETICAL PROTEIN; MEMBRANE; LIPOPROTEIN; SIGNAL.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 91 HYPOTHETICAL LIPOPROTEIN YPDI.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 91 AA; 10162 MW; 294E168F CRC32;

Query Match 94.9%; Score 37; DB 1; Length 91;

Best Local Similarity 83.3%; Pred. No. 1.81e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 48 HSKLQ 53
QY 1 HSKLQ 6

RESULT 6
ID GLP3_GLYDI STANDARD; PRT; 147 AA.
AC P21660;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)
DE GLOBIN, POLYMERIC COMPONENT P3.
OS GLYCERA DIBRANCHIATA (BLOODWORM).
OC EUKARYOTA; METAZOA; ANNELIDA; POLYCHAETA; PHYLLODOCIDA; GLYCERIDAE;
OC GLYCERA.
CC [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 91072390.
RA ZAFAR R.S., CHOW L.H., STERN M.S., SCULLY J.S., SHARMA P.R.,
RA VINOGRADOV S.N., WALZ D.A.;
RT "the cDNA sequences encoding two components of the polymeric fraction
RT of the intracellular hemoglobin of Glycera dibranchiata";
RL J. BIOL. CHEM. 265:21843-21851(1990).
CC -1- SUBUNIT: POLYMER.
CC -1- THIS PROTEIN IS ONE OF AT LEAST SIX COMPONENTS IN THE POLYMERIC
CC FRACTION OF GLYCERA HEMOGLOBIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55444; G159147;
DR PIR; B36529; B36529.
DR PROSITE; PS01033; GLOBIN; 1.
DR PFAM; PF00042; globin; 1.
DR HSP; P02216; IHG
KW HEME; OXYGEN TRANSPORT; RESPIRATORY PROTEIN.
SQ SEQUENCE 147 AA; 15977 MW; 747D26B7 CRC32;

Query Match 92.3%; Score 36; DB 1; Length 147;
Best Local Similarity 83.3%; Pred. No. 3.80e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 57 HSKLQ 62
QY 1 HSKLQ 6

RESULT 7
ID FMO3_RABIT STANDARD; PRT; 530 AA.
AC P32417;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE DIMETHYLANILINE MONOOXYGENASE [N-OXIDE FORMING] 3 (EC 1.14.13.8)
DE (HEPATIC FLAVIN-CONTAINING MONOOXYGENASE 3) (FMO 3) (DIMETHYLANILINE
DE OXIDASE 3) (FMO 1D1) (FMO FORM 2).
GN FMO3.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
CC [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=LIVER;
RX MEDLINE; 94245761.
RA BURNETT V.L., LAWTON M.P., PHILPOT R.M.;

"Cloning and sequencing of flavin-containing monooxygenases FMO3 and FMO4 from rabbit and characterization of FMO3.";
J. BIOL. CHEM. 269:14314-14322(1994).
[2]
RN SEQUENCE. 91378577.
RX MEDLINE. 91378577.
RA OZOLS J.;
RT "Multiple forms of liver microsomal flavin-containing monooxygenases: complete covalent structure of form 2.";
RL ARCH. BIOCHEM. BIOPHYS. 290:103-115(1991).
RN [3]
RP SEQUENCE OF 1-32.
RX TISSUE=LIVER;
RC MEDLINE; 89374273.
RA OZOLS J.;
RT "Liver microsomes contain two distinct NADPH-Monooxygenases with NH2-terminal segments homologous to the flavin containing NADPH-monooxygenase of Pseudomonas fluorescens";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 163:49-55(1989).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE OXIDATIVE METABOLISM OF A VARIETY OF XENOBIOTICS SUCH AS DRUGS AND PESTICIDES. FORM II N-OXYGENATES PRIMARY ALIPHATIC ALKYLAMINES AS WELL AS SECONDARY AND TERTIARY AMINES.
CC -!- CATALYTIC ACTIVITY: N,N-DIMETHYLANILINE + NADPH + O(2) = N,N-DIMETHYLANILINE N-OXIDE + NADP(+) + H(2)O.
CC -!- COFACTOR: FAD FLAVOPROTEIN.
CC -!- SUBCELLULAR LOCATION: MICROSOMAL.
CC -!- TISSUE SPECIFICITY: LIVER.
CC -!- SIMILARITY: BELONGS TO THE FMO FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; LI0391; G349534; -.
DR PIR; S18380; S18380.
DR PFAM; PF00743; FMO-like; 1.
KW OXIDOREDUCTASE; MONOOXYGENASE; NADP; FLAVOPROTEIN; FAD; MICROSOME;
KW TRANSMEMBRANE; MULTIGENE FAMILY.
FT INIT_MET 0 0
FT NP_BIND 8 13 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 190 195 NADP (POTENTIAL).
FT VARIANT 278 278 R -> M.
FT VARIANT 404 404 M -> V.
FT VARIANT 75 75 D -> P (IN REF. 2).
FT CONFLICT 80 80 F -> N (IN REF. 2).
FT CONFLICT 127 129 STE -> ATC (IN REF. 2).
FT CONFLICT 171 172 HS -> RQ (IN REF. 2).
FT CONFLICT 196 196 C -> E (IN REF. 2).
FT CONFLICT 305 305 F -> FKEF (IN REF. 2).
FT CONFLICT 418 418 W -> T (IN REF. 2).
FT CONFLICT 422 422 S -> W (IN REF. 2).
FT CONFLICT 513 515 WLK -> ELW (IN REF. 2).
SQ SEQUENCE 530 AA; 59684 MW; 8A14B159 CRC32;
Query Match 92.3%; Score 36; DB 1; Length 530;
Best Local Similarity 83.3%; Pred. No. 3.80e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 82 HNSKLQ 87
QY 1 HSSKLQ 6
RESULT 8
ID DRP2_HUMAN STANDARD; PRT; 954 AA.
AC Q13474;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DYSTROPHIN-RELATED PROTEIN 2.
GN DRP2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-CELL;
RX MEDLINE; 96225452.
RA ROBERTS R.G., FREEMAN T.C., KENDALL E., VETRIE D.L.P., DIXON A.K., SHAW-SMITH C., BONE O., BOBROW M.;
RT "Characterization of DRP2, a novel human dystrophin homologue.";
RL NAT. GENET. 13:223-226(1996).
CC -!- FUNCTION: POSSIBLY INVOLVED IN MEMBRANE-CYTOSKELETON INTERACTIONS OF THE CENTRAL NERVOUS SYSTEM.
CC -!- TISSUE SPECIFICITY: BRAIN AND SPINAL CORD; VERY LOW LEVELS IN EYE, OVARY EPIDIDYMIS AND TESTIS; NOT IN CARDIAC AND SKELETAL MUSCLE.
CC -!- SIMILARITY: IN OVERALL STRUCTURE TO DP116 DYSTROPHIN ISOFORM.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; U43519; G1353782; -.
DR MIM; 300052; -.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS50020; WW_DOMAIN_2; 1.
DR PFAM; PF00397; WW_rsp5_WWP; 1.
DR PFAM; PF00435; Spectrin; 2.
DR PFAM; PF00569; Z2; 1.
KW STRUCTURAL PROTEIN; CYTOSKELETON.
FT CA_BIND 20 31 SITE 1.
FT REPEAT 98 226 1.
FT REPEAT 228 348 2.
FT DOMAIN 355 380 WW DOMAIN.
FT DOMAIN 601 648 ZZ DOMAIN.
FT SIMILAR 424 451 TO EF-HAND.
FT SIMILAR 472 500 TO EF-HAND.
SQ SEQUENCE 954 AA; 107673 MW; 0197AA02 CRC32;
Query Match 92.3%; Score 36; DB 1; Length 954;
Best Local Similarity 83.3%; Pred. No. 3.80e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 100 HSGKLQ 105
QY 1 HSKKLQ 6
RESULT 9
ID BMB_Emeni STANDARD; PRT; 2067 AA.
AC P33144;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE CELL DIVISION-ASSOCIATED PROTEIN BMB.
GN BMB.
OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC EUROTIALES; TRICHOCOMACEAE; EMERICELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A773;
RX MEDLINE; 92348436.
RA MAY G.S., MCGOLDRICK C.A., HOLT C.L., DENISON S.H.;
RT "The bmb3 mutation of Aspergillus nidulans uncouples DNA replication from the completion of mitosis.";
RL J. BIOL. CHEM. 267:15737-15743(1992).

CC -!- FUNCTION: REQUIRED FOR NUCLEAR DIVISION. COULD FUNCTION IN THE
 CC MITOTIC SPINDLE.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- SIMILARITY: TO YEAST ESP1 AND S.POMBE CUT1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M83232; G168021; ALT_TERM.
 CC PIR: A42854; A42854.
 CC NUCLEAR PROTEIN: MITOSIS.
 CC
 CC SEQUENCE 2067 AA; 227925 MW; D3B877CC CRC32;

Query Match 89.7%; Score 35; DB 1; Length 2067;

Best Local Similarity 66.7%; Pred. No. 7.83e+00;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1665 HAAKIQ 1670

Qy 1 HSSKIQ 6

RESULT 10
 ID GLP2_GLYDI STANDARD; PRT; 147 AA.
 AC P21659;
 DT 01-MAY-1991 (REL. 18, CREATED)
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE GLOBIN, POLYMERIC COMPONENT P2.
 OS GLYCERA DIBRANCHIATA (BLOODWORM).
 OC EUKARYOTA; METAZOA; ANNELIDA; POLYCHAETA; PHYLLODOCIDA; GLYCERIDAE;
 CC GLYCERA.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE; 91072390.
 CC ZAFAR R.S., CHOW L.H., STERN M.S., SCULLY J.S., SHARMA P.R.,
 CC VINOGRADOV S.N., WALZ D.A.;
 CC "The cDNA sequences encoding two components of the polymeric fraction
 CC of the intracellular hemoglobin of the glycera dibranchiata";
 CC J. BIOL. CHEM. 265:21843-21851(1990).
 CC -!- SUBUNIT: POLYMER.
 CC -!- THIS PROTEIN IS ONE OF AT LEAST SIX COMPONENTS IN THE POLYMERIC
 CC FRACTION OF GLYCERA HEMOGLOBIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M55443; G159145;
 CC PIR: A36529; A36529.
 CC PROSITE; PS01033; GLOBIN; 1.
 CC PFAM; PF00042; globin; 1.
 CC HSP; P02216; 1HGB.
 CC HEME; OXYGEN TRANSPORT; RESPIRATORY PROTEIN.
 CC SEQUENCE 147 AA; 15904 MW; ACA841AB CRC32;

Query Match 87.2%; Score 34; DB 1; Length 147;

Best Local Similarity 66.7%; Pred. No. 1.58e+01;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 57 HNAKIQ 62

Qy 1 HSSKIQ 6

RESULT 11
 ID TRMU_SCHPO STANDARD; PRT; 415 AA.
 AC Q13947;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PROBABLE TRNA (5-METHYLAMINOMETHYL-2-THIOURIDYLATE)-METHYLTRANSFERASE
 DE (EC 2.1.1.61).
 GN SPAC23H4.04.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIZOSOMYCETES;
 CC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 CC SCHIZOSACCHAROMYCES.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=972;
 CC BROWN D., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 CC SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA = S-ADENOSYL-L-
 CC HOMOCYSTEINE + TRNA CONTAINING 5-METHYLAMINOMETHYL-2-
 CC THIOURIDYLATE.
 CC -!- SIMILARITY: BELONGS TO THE TRMU FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z98977; E339919;
 CC TRANSFERASE; METHYLTRANSFERASE; TRNA PROCESSING.
 CC SEQUENCE 415 AA; 47626 MW; 36B7C485 CRC32;

Query Match 87.2%; Score 34; DB 1; Length 415;

Best Local Similarity 66.7%; Pred. No. 1.58e+01;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 21 HSTRLO 26

Qy 1 HSSKIQ 6

RESULT 12
 ID Y187_MYCGE STANDARD; PRT; 585 AA.
 AC P47433;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN MG187.
 GN MG187.
 OS MYCOPLASMA GENITALIUM.
 CC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLIICUTES;
 CC MYCOPLASMATACEAE; MYCOPLASMA.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=ATCC 33530 / G-37;
 CC MEDLINE; 96026346.
 CC FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
 CC FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
 CC FRITCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
 CC NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
 CC TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
 CC PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
 CC "The minimal gene complement of Mycoplasma genitalium";
 CC SCIENCE 270:397-403(1995).
 CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; U39696; G1045872; -
DR TIGR; MG187; -
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_tran; 1
KW HYPOTHETICAL PROTEIN; ATP-BINDING; TRANSPORT.
FT NP_BIND 40 47 ATP (POTENTIAL).
SQ SEQUENCE 585 AA; 66949 MW; EC78C4B8 CRC32;

Query Match 87.2%; Score 34; DB 1; Length 585;
Best Local Similarity 66.7%; Pred. No. 1.58e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 226 HNAKLQ 231
|::|||
Qy 1 HSSKLQ 6

RESULT 13
ID CID_DROME STANDARD; PRT; 1377 AA.
AC P19538;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE CUBITUS INTERRUPTUS DOMINANT PROTEIN.
GN CI-D.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE; 90346286.
RA ORENIC T.V., SLUSARSKI D.C., KROLL K.L., HOLMGREN R.A.;
RT "Cloning and characterization of the segment polarity gene cubitus
interruptus dominant of Drosophila.";
RL GENES DEV. 4:1053-1067(1990).

CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=OREGON-R;
CC MEDLINE; 92146935.
RA BERRY A.J., AJIOKA J.W., KREITMAN M.;
RT "Lack of polymorphism on the Drosophila fourth chromosome resulting
RT from selection.";
RL GENETICS 129:1111-1117(1991).
CC -!- FUNCTION: INVOLVED IN SEGMENT POLARITY. IS REQUIRED FOR THE NORMAL
CC DEVELOPMENT OF THE POSTERIOR HALF OF EACH EMBRYONIC SEGMENT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED UNIFORMLY THROUGHOUT THE BLASTODERM
CC STAGE AND GASTRULATION AND DOES NOT RESOLVE INTO SEGMENTALLY
CC REPEATING STRIPES UNTIL THE END OF THE SHORT PHASE OF GERM-BAND
CC EXTENSION.
CC -!- SIMILARITY: TO THE GLI-RELATED GROUP OF C2H2-TYPE ZINC-FINGERS
CC PROTEINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X54360; G7733; -
CC PIR; A35817; A35817.
CC PIR; S12769; S12769.
CC FLYBASE; FBgn0004859; ci.

DR PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
DR PFAM; PF00096; zf-C2H2; 5.
DR HSSP; P08151; 2GLI.
KW DEVELOPMENTAL PROTEIN; SEGMENTATION POLARITY PROTEIN; ZINC-FINGER;
KW METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
FT DOMAIN 451 603 ZINC-FINGERS.
FT ZN_FING 451 476 C2H2-TYPE.
FT ZN_FING 484 511 C2H2-TYPE.
FT ZN_FING 517 541 C2H2-TYPE.
FT ZN_FING 547 572 C2H2-TYPE.
FT ZN_FING 578 603 C2H2-TYPE.
SQ SEQUENCE 1377 AA; 150881 MW; A14EB3FC CRC32;

Query Match 87.2%; Score 34; DB 1; Length 1377;
Best Local Similarity 66.7%; Pred. No. 1.58e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 271 HSTRLO 276
|::|||
Qy 1 HSSKLQ 6

RESULT 14
ID Y133_HUMAN STANDARD; PRT; 1524 AA.
AC Q14146;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN KIAA0133.
GN KIAA0133
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE; 96127530.
RA NAGASE T., SEKI N., TANAKA A., ISHIKAWA K.-I., NOMURA N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA RES. 2:167-174(1995).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D50923; G1469189; -
CC HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
CC TRANSMEM 119 139 POTENTIAL.
CC TRANSMEM 164 184 POTENTIAL.
CC TRANSMEM 292 312 POTENTIAL.
CC TRANSMEM 533 553 POTENTIAL.
CC TRANSMEM 631 651 POTENTIAL.
CC TRANSMEM 725 745 POTENTIAL.
CC TRANSMEM 747 767 POTENTIAL.
CC TRANSMEM 906 926 POTENTIAL.
CC TRANSMEM 993 1013 POTENTIAL.
CC TRANSMEM 1150 1170 POTENTIAL.
CC TRANSMEM 1310 1330 POTENTIAL.
CC TRANSMEM 1335 1355 POTENTIAL.
CC TRANSMEM 1462 1482 POTENTIAL.
CC SEQUENCE 1524 AA; 170569 MW; E074D413 CRC32;

Query Match 87.2%; Score 34; DB 1; Length 1524;
Best Local Similarity 83.3%; Pred. No. 1.58e+01;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Search completed: Thu Oct 28 11:20:45 1999
Job time : 8 secs.

```
Db 79 HSKLQ 84
   || |||
QY 1 HSKLQ 6

RESULT 15
ID MYSJ_DICDI STANDARD; PRT; 2245 AA.
AC P54697;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE MYOSIN IJ HEAVY CHAIN.
GN MYOJ.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX3;
RA HAMMER J.A. III, JUNG G.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RA PETERSON M.D., TITUS M.A.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE; 95023928.
RA TITUS M.A., KUSPA A., LOOMIS W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:9446-9450(1994).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO CLASS-5 MYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U42409; G1150766; -
DR EMBL; L35322; G1039361; -
DR DICTYDB; DD01095; MYOJ.
DR PFAM; PF00063; myosin_head; 2.
DR PFAM; PF00612; IQ; 3.
DR HSP; P08799; 1MND.
KW MYOSIN; REPEAT; ATP-BINDING; CALMODULIN-BINDING; ACTIN-BINDING;
KW HEPTAD REPEAT PATTERN; COILED COIL.
FT DOMAIN 1 ? ?
FT DOMAIN ? ?
FT DOMAIN 959 1812 COILED COIL (POTENTIAL).
FT DOMAIN ? 2245 TAIL.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 669 749 ACTIN-BINDING.
FT CONFLICT 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKSGCFEIEGVSDDEH -> IENMFELKVRMKS (IN
FT REF. 2).
FT CONFLICT 550 550 N -> K (IN REF. 2).
FT CONFLICT 865 866 HH -> QQ (IN REF. 2).
SQ SEQUENCE 2245 AA; 258478 MW; 48F462D8 CRC32;

Query Match 87.2%; Score 34; DB 1; Length 2245;
Best local Similarity 83.3%; Pred. No. 1.58e+01;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1247 HSKLQ 1252
   || |||
QY 1 HSKLQ 6
```

THIS PAGE BLANK (USPTO)

WASREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:21:27 1999; Maspar time 4.18 Seconds
Tabular output not generated. 78.265 Million cell updates/sec

Title: >US-09-081-707-7
Description: (1-6) from US09081707.pep
Perfect Score: 39
Sequence: 1 HSSKQLQ 6

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sp_trembl9

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 17.599; Variance 13.927; scale 1.264

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	37	94.9	278	2 Q46797	WAS F268 AND F157P BEF	4.33e+00
2	37	94.9	378	8 O20404	CYTCHROME B.	4.33e+00
3	37	94.9	378	8 O21145	CYTCHROME B.	4.33e+00
4	37	94.9	883	10 O23511	RNA HELICASE.	4.33e+00
5	36	92.3	509	10 O81317	F6N15.5 PROTEIN.	8.73e+00
6	36	92.3	544	5 Q17060	ROYAL JELLY PROTEIN RJ	8.73e+00
7	36	92.3	627	4 P78335	GUANYLATE KINASE ASSOC	8.73e+00
8	36	92.3	666	11 P97841	GUANYLATE KINASE ASSOC	8.73e+00
9	36	92.3	675	4 O14489	DAP-1 BETA.	8.73e+00
10	36	92.3	592	11 O54773	PSD-95 BINDING PROTEIN	8.73e+00
11	36	92.3	744	14 Q82039	OUTER CAPSID SPIKE PRO	8.73e+00
12	36	92.3	744	14 Q82040	OUTER CAPSID SPIKE PRO	8.73e+00
13	36	92.3	977	4 O14490	DAP-1 ALPHA.	8.73e+00
14	36	92.3	992	11 P97836	PSD-95/SAP90-ASSOCIATE	8.73e+00
15	36	92.3	1548	5 P91339	SIMILARITY TO HUMAN PR	8.73e+00
16	36	92.3	1638	5 O61529	GUANINE NUCLEOTIDE EXC	8.73e+00
17	36	92.3	2488	5 O61528	GUANINE NUCLEOTIDE EXC	8.73e+00
18	35	89.7	539	2 O85280	50KDA STRAIN-SPECIFIC	1.74e+01
19	35	89.7	764	5 Q24708	CORKSCREW PROTEIN (EC	1.74e+01
20	35	89.7	836	5 Q17430	HYPOTHETICAL 95.5 KD P	1.74e+01

21	35	89.7	849	2 O85281	85KDA STRAIN-SPECIFIC	1.74e+01
22	35	89.7	1087	10 O82303	PUTATIVE RNA HELICASE.	1.74e+01
23	35	89.7	1167	10 O49471	TMV RESISTANCE PROTEIN	3.40e+01
24	34	87.2	105	10 Q40128	FLOWER-SPECIFIC GAMMA-	3.40e+01
25	34	87.2	225	4 Q13687	OEC/ABL FUSION PROTEIN	3.40e+01
26	34	87.2	265	10 Q40407	OEC 23KD PROTEIN.	3.40e+01
27	34	87.2	284	4 Q13692	BCR/ABL FUSION PROTEIN	3.40e+01
28	34	87.2	294	5 Q24847	SURFACE ANTIGEN.	3.40e+01
29	34	87.2	332	5 Q22982	COSMID F47D2.	3.40e+01
30	34	87.2	492	14 O89808	ECDYSTEROID UDP-GLUCOS	3.40e+01
31	34	87.2	579	10 O80953	T16B24.8 PROTEIN.	3.40e+01
32	34	87.2	1146	6 O46490	BRCA1 (FRAGMENT).	3.40e+01
33	34	87.2	1146	6 O46488	BRCA1 (FRAGMENT).	3.40e+01
34	34	87.2	1397	5 O18525	CUBITUS INTERRUPTUS DO	3.40e+01
35	34	87.2	1795	2 O68101	AVIRULENCE PROTEIN.	3.40e+01
36	34	87.2	2823	5 O45614	T22A3.8 PROTEIN.(FRAG	3.40e+01
37	33	84.6	209	2 O87949	TORA SPECIFIC CHAPERON	6.55e+01
38	33	84.6	210	2 P70747	MAJOR SURFACE PROTEIN	6.55e+01
39	33	84.6	211	2 O84703	ENDONUCLEASE III.	6.55e+01
40	33	84.6	305	10 O81659	ISOPENTENYL PYROPHOSPH	6.55e+01
41	33	84.6	341	8 O09430	CYTCHROME OXIDASE I (6.55e+01
42	33	84.6	420	2 Q51709	CYTCHROME BA(3) PRECU	6.55e+01
43	33	84.6	753	4 Q09471	M-PHASE PHOSPHOPROTEIN	6.55e+01
44	33	84.6	754	13 O13131	SIGNAL TRANSDUCER/ACTI	6.55e+01
45	33	84.6	896	14 Q82232	PUTATIVE FRAMESHIFTING	6.55e+01

ALIGNMENTS

RESULT 1
ID Q46797 PRELIMINARY; PRT; 278 AA.
AC Q46797;
DT 01-NOV-1996 (TREMUREL. 01, CREATED)
DT 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMUREL. 02, LAST ANNOTATION UPDATE)
DE WAS F268 AND F157P BEFORE SPLICE.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RC STRAIN-K-12;
RA PLUNKETT G.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T.,
RA GLASNER F.D.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U28375; G887814; -;
DR EMBL; AE000369; G1789227; -;
SQ SEQUENCE 278 AA; 31319 MW; ECF6C510 CRC32;

Query Match 94.9%; Score 37; DB 2; Length 278;
Best Local Similarity 83.3%; Pred. No. 4.33e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 77 HSAKLO 82
||:||||
Qy 1 HSSKLO 6

RESULT 2
ID O20404 PRELIMINARY; PRT; 378 AA.
AC O20404;
DT 01-JAN-1998 (TREMUREL. 05, CREATED)
DT 01-JAN-1998 (TREMUREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMUREL. 08, LAST ANNOTATION UPDATE)
DE CYTOCHROME B.
GN CYTB.
OS TOMOCICHLA TUBA.
OG MITOCHONDRION.

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
 OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
 OC LABROIDEI; CICHLIDAE; TOMOCICHLA.
 RN [1]
 RA MARTIN A.P., BIRMINGHAM E.;
 RP SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
 CC FERROCYTOCHROME C.
 CC -1- COFACTOR: TWO HEME GROUPS
 CC (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
 CC (BY SIMILARITY).
 DR EMBL: AF009941; G2281563; -
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 DR PFAM: PF00032; cytochrome_b_c; 1.
 DR PFAM: PF00033; cytochrome_b_n; 1.
 DR PFAM: PF00033; cytochrome_b_n; 1.
 KW MITOCHONDRION; ELECTRON TRANSPORT; RESPIRATORY CHAIN; TRANSMEMBRANE;
 KW HEME.
 SQ SEQUENCE 378 AA; 42344 MW; 2E7A3348 CRC32;
 Query Match 94.9%; Score 37; DB 8; Length 378;
 Best Local Similarity 83.3%; Pred. No. 4.33e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 308 HTSKLQ 313
 QY 1 HSSKLQ 6
 RESULT 3
 ID Q21145 PRELIMINARY; PRT; 378 AA.
 AC Q21145
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CYTOCHROME B.
 GN CYTB.
 OS CAQUETAIA UMBIFERUM.
 OC MITOCHONDRION.
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
 OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
 OC LABROIDEI; CICHLIDAE; CAQUETAIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MARTIN A.P., BIRMINGHAM E.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
 CC FERROCYTOCHROME C.
 CC -1- COFACTOR: TWO HEME GROUPS
 CC (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
 CC (BY SIMILARITY).
 DR EMBL: AF009940; G2281561; -
 DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 DR PFAM: PF00032; cytochrome_b_c; 1.
 DR PFAM: PF00033; cytochrome_b_n; 1.
 DR PFAM: PF00033; cytochrome_b_n; 1.
 KW MITOCHONDRION; ELECTRON TRANSPORT; RESPIRATORY CHAIN; TRANSMEMBRANE;
 KW HEME.
 SQ SEQUENCE 378 AA; 42425 MW; E7F49DA3 CRC32;
 Query Match 94.9%; Score 37; DB 8; Length 378;
 Best Local Similarity 83.3%; Pred. No. 4.33e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 308 HTSKLQ 313
 QY 1 HSSKLQ 6
 RESULT 4
 ID Q23511 PRELIMINARY; PRT; 883 AA.
 AC Q23511
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE RNA HELICASE.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98121113.
 RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
 RA BERKAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
 RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFFANELLI P., WEDLER H.,
 RA WEDLER E., WAMBUIT R., WEITZENEGGER T., POHL T.M., TERRY N.,
 RA GIELEN J., VILLARROEL R., DE CLERCQ R., VAN MONTAGU M., LECHARNY A.,
 RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,
 RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
 RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
 RA VOUKELATOU E., MILLIONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,
 RA HILBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
 RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
 RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
 RA SCHUELLER C., CHALWATZIS N.;
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 RT Arabidopsis thaliana.";
 RL NATURE 391:485-488(1998).
 DR EMBL: Z97341; E327007; -
 DR PROSITE: PS00690; DEAH_ATP_HELICASE; 1.
 DR PFAM: PF00271; helicase_C; 1.
 KW HELICASE.
 SQ SEQUENCE 883 AA; 100986 MW; 3811F298 CRC32;
 Query Match 94.9%; Score 37; DB 10; Length 883;
 Best Local Similarity 83.3%; Pred. No. 4.33e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 771 HSAKLQ 776
 QY 1 HSSKLQ 6
 RESULT 5
 ID Q81317 PRELIMINARY; PRT; 509 AA.
 AC Q81317
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE F6N15.5 PROTEIN.
 GN F6N15.5.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA WASHU;
 RT "The A. thaliana Genome Sequencing Project.";
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA RYAN E., EDWARDS J., PAPE K.;
 RT "The sequence of A. thaliana F6N15.";
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA WATERSTON R.;
 RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF069299; G3193321; -
SQ SEQUENCE 509 AA; 58562 MW; 5A682D0D CRC32;

Query Match 92.3%; Score 36; DB 10; Length 509;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 376 HSPKIQ 381
||:||||
QY 1 HSSKIQ 6

RESULT 6
ID Q17060 PRELIMINARY; PRT; 544 AA.
AC Q17060;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE. ROYAL JELLY PROTEIN RJP57-1 PRECURSOR.
OS APIS MELLIFERA (HONEYBEE).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; HYMENOPTERA; APOCRITA; ACULEATA; APOIDEA; APIDAE; APIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8 DAYS OLD NURSE HONEY BEE; TISSUE=HEAD;
RA KLAUDINY J., HANES J., KULIFAJOVA J., ALBERT S., SIMUTH J.;
RL J. APICULTURAL RES. 33:105-111(1994).
DR EMBL; 226318; E213552; -
KW SIGNAL.
FT SIGNAL
SQ SEQUENCE 544 AA; 61662 MW; 767820E6 CRC32;

Query Match 92.3%; Score 36; DB 5; Length 544;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 539 HSSKLH 544
||||:
QY 1 HSSKIQ 6

RESULT 7
ID P78335 PRELIMINARY; PRT; 627 AA.
AC P78335;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE GUANYLATE KINASE ASSOCIATED PROTEIN.
GN GRAP.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97177144.
RA KIM E., NAISBITT S., HSUEH Y.P., RAO A., ROTHSCCHILD A., CRAIG A.M.,
RA SHENG M.;
RT "GRAP, a novel synaptic protein that interacts with the guanylate
RT kinase-like domain of the PSD-95/SAP90 family of channel clustering
RT molecules.";
RL J. CELL BIOL. 136:669-678(1997).
DR EMBL; U67988; G1857137; -
SQ SEQUENCE 627 AA; 70051 MW; C40A3CBB CRC32;

Query Match 92.3%; Score 36; DB 4; Length 627;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 94 HSPKIQ 99
||:||||
QY 1 HSSKIQ 6

RESULT 8
ID P97841 PRELIMINARY; PRT; 666 AA.
AC P97841;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GUANYLATE KINASE ASSOCIATED PROTEIN.
GN GRAP.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97177144.
RA KIM E., NAISBITT S., HSUEH Y.P., RAO A., ROTHSCCHILD A., CRAIG A.M.,
RA SHENG M.;
RT "GRAP, a novel synaptic protein that interacts with the guanylate
RT kinase-like domain of the PSD-95/SAP90 family of channel clustering
RT molecules.";
RL J. CELL BIOL. 136:669-678(1997).
DR EMBL; U67987; G1857139; -
SQ SEQUENCE 666 AA; 74233 MW; B3E3BB93 CRC32;

Query Match 92.3%; Score 36; DB 11; Length 666;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 104 HSPKIQ 109
||:||||
QY 1 HSSKIQ 6

RESULT 9
ID O14489 PRELIMINARY; PRT; 675 AA.
AC O14489;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE DAP-1 BETA.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97431353.
RA SATOH K., YANAI H., SENDA T., KOHU K., NAKAMURA T., OKUMURA N.,
RA MATSUMINE A., KOBAYASHI S., TOYOSHIMA K., AKIYAMA T.;
RT "DAP-1, a novel protein that interacts with the guanylate kinase-like
RT domains of hDLG and PSD-95.";
RL GENES TO CELLS 2:415-424(1997).
DR EMBL; AB000276; D1024134; -
SQ SEQUENCE 675 AA; 75260 MW; 1FFE2AC4 CRC32;

Query Match 92.3%; Score 36; DB 4; Length 675;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 94 HSPKIQ 99
||:||||
QY 1 HSSKIQ 6

RESULT 10
ID O54773 PRELIMINARY; PRT; 692 AA.
AC O54773;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PSD-95 BINDING PROTEIN.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE: 98089008.
RA KAWASHIMA N., TAKAMIYA K., SUN J., KITABATAKE A., SOBUE K.;
RT "Differential expression of isoforms of PSD-95 binding protein
RL (GKAP/SAPAP1) during rat brain development.";
RL FEBS LETT 418:301-304(1997).
DR EMBL: AB003594; D1025176; -.
SQ SEQUENCE 692 AA; 76991 MW; CD061633 CRC32;

Query Match 92.3%; Score 36; DB 11; Length 692;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 102 HSPKQ 107
||:|
QY 1 HSSKQ 6

RESULT 11
ID Q82039 PRELIMINARY; PRT; 744 AA.

AC Q82039;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE OUTER CAPSID SPIKE PROTEIN.
GN VP4.

OS HUMAN ROTAVIRUS.
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BELEM;

RX MEDLINE: 94378523.
RA FIELDING P.A., LAMB DEN P.R., CAUL E.O., CLARKE I.N.;

RT "Molecular characterization of the outer capsid spike protein (VP4)
RT gene from human group C rotavirus.";
RL VIROLOGY 204:442-446(1994).
DR EMBL: X79441; G495229; -.

DR PFAM: PF00426; VP4; 1.
SQ SEQUENCE 744 AA; 84036 MW; CDD47684 CRC32;

Query Match 92.3%; Score 36; DB 14; Length 744;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 410 HSSRLQ 415
||:|
QY 1 HSSKQ 6

RESULT 12
ID Q82040 PRELIMINARY; PRT; 744 AA.

AC Q82040;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE OUTER CAPSID SPIKE PROTEIN.
GN VP4.

OS HUMAN ROTAVIRUS.
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL;

RX MEDLINE: 94378523.
RA FIELDING P.A., LAMB DEN P.R., CAUL E.O., CLARKE I.N.;

RT "Molecular characterization of the outer capsid spike protein (VP4)
RT gene from human group C rotavirus.";
RL VIROLOGY 204:442-446(1994).
DR EMBL: X79442; G495231; -.

DR PFAM: PF00426; VP4; 1.
SQ SEQUENCE 744 AA; 84080 MW; 4E128877 CRC32;

Query Match 92.3%; Score 36; DB 14; Length 744;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 410 HSSRLQ 415
||:|
QY 1 HSSKQ 6

RESULT 13
ID O14490 PRELIMINARY; PRT; 977 AA.

AC O14490;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE DAP-1 ALPHA.

OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97431353.

RA SATOH K., YANAI H., SENDA T., KOHU K., NAKAMURA T., OKUMURA N.,
RA MATSUMINE A., KOBAYASHI S., TOYOSHIMA K., AKIYAMA T.;

RT "DAP-1, a novel protein that interacts with the guanylate kinase-like
RT domains of hdlg and PSD-95.";
RL GENES TO CELLS 2:415-424(1997).

DR EMBL: AB000277; D1024135; -.

SQ SEQUENCE 977 AA; 108872 MW; FFF0A0A0 CRC32;

Query Match 92.3%; Score 36; DB 4; Length 977;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 396 HSPKQ 401
||:|
QY 1 HSSKQ 6

RESULT 14
ID P97836 PRELIMINARY; PRT; 992 AA.

AC P97836;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PSD-95/SAP90-ASSOCIATED PROTEIN-1.

OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.

RN [1]
RP SEQUENCE FROM N.A.
RA TAKAI Y., TAKEUCHI M., IRIE M., HATA Y.;

RL J. BIOL. CHEM. 0:0-0(0).
DR EMBL: U67137; G1864087; -.

SQ SEQUENCE 992 AA; 110177 MW; 7D090FD0 CRC32;

Query Match 92.3%; Score 36; DB 11; Length 992;
Best Local Similarity 83.3%; Pred. No. 8.73e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 402 HSPKQ 407
||:|
QY 1 HSSKQ 6

RESULT 15
ID P91339 PRELIMINARY; PRT; 1548 AA.

AC P91339;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILARITY TO HUMAN PROTO-ONCOGENE DBL.
GN F55C7.7.

OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HANKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALLDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA DU Z., LE T.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U80436; GI703564; -.
 DR PFAM; PF00621; RhogEF; 1.
 SQ SEQUENCE 1548 AA; 179071 MW; 63FBFA8A CRC32;

Query Match 92.3%; Score 36; DB 5; Length 1548;
 Best Local Similarity 83.3%; Pred. No. 8.73e+00;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 437 HSSRLQ 442
 |||:|
 QY 1 HSSKLQ 6

Search completed: Thu Oct 28 11:22:09 1999
 Job time : 42 secs.

THIS PAGE BLANK (USPTO)

W P S R L
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Oct 28 11:24:16 1999; Maspar time 3.30 Seconds
Tabular output not generated. 45.117 Million cell updates/sec

Title: >US-09-081-707-10
Description: (1-7) from US09081707.pgp
Perfect Score: 45
Sequence: 1 EHSSKIQ 7
Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 13.622; Variance 31.430; scale 0.433

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	42	93.3	992 34	W69742	SAPAP2 protein.	4.21e+01
2	42	93.3	992 34	W69740	SAPAP1 protein.	4.21e+01
3	38	84.4	210 35	W70284	Anaplasma marginale m	1.51e+02
4	38	84.4	462 26	W33420	Human semenogelin I.	1.51e+02
5	38	84.4	462 17	R99583	Human semenogelin I.	1.51e+02
6	36	80.0	432 36	W27182	HSV-2 strain SB5 Cont	2.80e+02
7	36	80.0	497 35	W27030	HSV-2 strain SB5 Cont	2.80e+02
8	36	80.0	497 36	W72132	HSV-2 strain SB5 Cont	2.80e+02
9	36	80.0	735 34	W69752	Herpes simplex virus-	2.80e+02
10	35	77.8	212 2	R08121	Modified tilapia prol	3.81e+02
11	35	77.8	212 2	R06893	Tilapia prolactin I.	3.81e+02
12	35	77.8	742 14	R74094	Human zona pellucida-	3.81e+02
13	35	77.8	3431 5	R25138	SFV4 non-structural p	3.81e+02
14	35	77.8	3084 35	W50891	Mouse laminin A chain	3.81e+02
15	34	75.6	7 24	W27419	CDR2 from light chain	5.15e+02
16	34	75.6	118 18	R97559	Tomato pZ130 cDNA-enc	5.15e+02

17	34	75.6	118 20	W08364	Tomato ovary-specific	5.15e+02
18	34	75.6	127 24	W21852	Humanised light chain	5.15e+02
19	34	75.6	127 24	W21855	Humanised light chain	5.15e+02
20	34	75.6	127 24	W21853	Humanised light chain	5.15e+02
21	34	75.6	145 29	W38463	Mouse RNA-binding pro	5.15e+02
22	34	75.6	179 2	R10310	Ovary tissue transcri	5.15e+02
23	34	75.6	184 29	W38464	Mouse RNA-binding pro	5.15e+02
24	34	75.6	188 6	R30779	pZ130 polyepitide.	5.15e+02
25	34	75.6	240 26	W36101	Mutant arylmalonate d	5.15e+02
26	34	75.6	409 17	R85868	WD-40 domain-contg. h	5.15e+02
27	34	75.6	410 13	R70005	OPDE 45 kDa subunit.	5.15e+02
28	34	75.6	439 29	W38466	Mouse RNA-binding pro	5.15e+02
29	34	75.6	506 34	W61152	Maltose binding prote	5.15e+02
30	34	75.6	524 27	W43304	Modified p17 antigen	5.15e+02
31	34	75.6	568 17	R96210	AIL protein-maltose b	5.15e+02
32	34	75.6	588 17	R96209	Invasin-maltose bindi	5.15e+02
33	34	75.6	659 21	W09294	L-proline-4-hydroxyla	5.15e+02
34	34	75.6	708 20	W06411	Maltose binding prote	5.15e+02
35	34	75.6	711 20	W06412	Maltose binding prote	5.15e+02
36	34	75.6	799 28	W23938	Wheat granule-bound s	5.15e+02
37	34	75.6	993 1	R05935	Secreted GPIIb subuni	5.15e+02
38	34	75.6	1084 31	W59033	B. pallidus DNA polym	5.15e+02
39	34	75.6	1181 23	W21727	Streptokinase/maltose	5.15e+02
40	34	75.6	1194 23	W21726	Streptokinase/maltose	5.15e+02
41	34	75.6	1194 23	W21725	Modified streptokinas	5.15e+02
42	34	75.6	1194 23	W21724	Modified streptokinas	5.15e+02
43	34	75.6	1276 31	W59034	B. pallidus DNA polym	5.15e+02
44	34	75.6	1607 35	W50897	Mouse laminin GI chal	5.15e+02
45	34	75.6	3084 2	P94758	Sequence of mouse lam	5.15e+02

ALIGNMENTS

RESULT 1
ID W69742 standard; protein; 992 AA.
AC W69742:
DT 26-OCT-1998 (first entry)
DE SAPAP2 protein.
KW Human; SAPAP2; SAPAP1; animal protein; PSD-95/SAP90; diagnosis;
KW nervous disease; functional interference; structural interference;
KW membrane associated guanylate kinase; neuronal disease.
OS Homo sapiens.
PN J10201478-A.
PD 04-AUG-1998.
PF 24-JAN-1997; 011715.
PR 24-JAN-1997; JP-011715.
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (TAKE/) TAKEUCHI M.
DR WPI; 98-474492/41.
PT DNA encoding new animal protein SAPAP 2 - useful for diagnosis and treatment of nervous system diseases
PS Claim 1; Page 7-9; 11pp; Japanese.
CC The present sequence represents the SAPAP2 protein. Also described in the present invention is: (A) an animal protein having an amino acid sequence substantially the same as SAPAP2; (B) a cDNA sequence encoding the amino acid sequence of SAPAP2 or (C) a genomic DNA sequence hybridised CC substantially same as SAPAP2; and (D) a genomic DNA sequence hybridised CC by the above cDNA or its partial sequence. SAPAP2 is a new animal CC protein which combines specifically with PSD-95/SAP90 and its related CC protein and is useful for the diagnosis, prevention and treatment of CC various nervous diseases caused by functional or structural interference CC of nervous system.
SQ Sequence 992 AA;

Query Match 93.3%; Score 42; DB 34; Length 992;
Best Local Similarity 85.7%; Pred.No. 4.21e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 401 ehspklq 407
111:111
QY 1 EHSSKIQ 7

RESULT 2
ID W69740 standard; protein; 992 AA.
AC W69740;
DT 26-OCT-1998 (first entry)
DE SAPAP1 protein.
KW Human; SAPAP1; SAPAP2; animal protein; PSD-95/SAP90; diagnosis;
KW nervous disease; functional interference; structural interference;
KW membrane associated guanylate kinase; neuronal disease.
OS Homo sapiens.
PN J10201477-A.
PD 04-AUG-1998.
PF 24-JAN-1997; 011714.
PR 24-JAN-1997; JP-011714.
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (TAKE/) TAKEUCHI M.
DR WPI; 98-474491/41.
PT New protein SAPAP1 - used for, e.g. diagnosis and prevention of
PT various neuronal diseases
PS Claim 1; Page 4-7; 13pp; Japanese.
CC The present sequence represents the SAPAP1 protein having a 992 amino
CC acid (aa) sequence. Also described in the present invention are: (1) an
CC animal protein having an aa sequence substantially homologous to SAPAP1;
CC (2) cDNA sequence encoding SAPAP1, or an aa sequence substantially
CC homologous to SAPAP1, and (3) a genomic DNA sequence hybridised to the
CC cDNA or its partial sequence. SAPAP1 is a novel animal protein specific
CC for PSD-95/SAP90 and its related protein, and may be useful for the
CC diagnosis, prevention and treatment of various neuronal diseases caused
CC by functional or structural interference of nervous system.
SQ Sequence 992 AA;

Query Match 93.3%; Score 42; DB 34; Length 992;
Best Local Similarity 85.7%; Pred. No. 4.21e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 401 ehspklq 407
QY 1 EHSSKLQ 7

RESULT 3
ID W70284 standard; protein; 210 AA.
AC W70284;
DT 18-NOV-1998 (first entry)
DE Anaplasma marginale major surface protein-5.
KW Anaplasma marginale major surface protein-5; MSP-5; anaplasmosis;
KW rickettsia; cattle; sheep; goat; ELISA; monoclonal antibody ANAF16C1;
KW mAb; competitive inhibition.
OS Anaplasma marginale.
PN US5798219-A.
PD 25-AUG-1998.
PF 16-OCT-1996; 730995.
PR 23-NOV-1993; US-156426.
PR 16-OCT-1996; US-730995.
PA (UNIV) UNIV WASHINGTON STATE RES FOUND.
PA (USDA) US SEC OF AGRIC.
PI Davis WC, Knowles DP, McElwain TF, McGuire TC, Palmer GH;
DR WPI; 98-480378/41.
PT ELISA detection of Anaplasma species in rickettsia - uses specific
PT monoclonal antibodies against Anaplasma major surface protein 5
PS Disclosure; Columns 5-8; 7pp; English.
CC The present sequence represents the Anaplasma marginale major surface
CC protein-5 (MSP-5). The invention provides a method for detecting
CC anaplasmosis, a vector-borne rickettsial disease of cattle, sheep and
CC goats. The method involves using the monoclonal antibody (mAb) ANAF16C1
CC and the Anaplasma MSP-5 protein in a competitive inhibition-ELISA format
CC for the serological detection of cattle, sheep or goats infected with
CC A. marginale. The test is specific for Anaplasma species as the MSP-5
CC protein detected is found in all Anaplasma species, and the mAb ANAF16C1
CC used to detect it is specifically raised against it. The detection
CC method is claimed to be capable of accurately detecting the protein due
CC to its specificity. Previous methods e.g. nucleic acid hybridisation,
CC cannot always detect carriers because of cyclic changes in rickettsia.
SQ Sequence 210 AA;

Query Match 84.4%; Score 38; DB 35; Length 210;
Best Local Similarity 57.1%; Pred. No. 1.51e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 102 dhadklq 108
QY 1 EHSSKLQ 7

RESULT 4
ID W33420 standard; protein; 462 AA.
AC W33420;
DT 27-MAR-1998 (first entry)
DE Human semenogelin I.
KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA; human;
KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
KW detection; cytotoxic conjugate; activation; quantitative assay.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Cleavage_site 159..160 /note= "Cleavage site 5"
FT Cleavage_site 289..290 /note= "Cleavage site 3"
FT Cleavage_site 316..317 /note= "Cleavage site 4"
FT Cleavage_site 349..350 /note= "Cleavage site 1"
FT Cleavage_site 376..377 /note= "Cleavage site 2"
FT W09712624-AI.
PD 10-APR-1997.
PF 02-OCT-1996; U15713.
PR 06-OCT-1995; US-540412.
PA (MERI) MERCK & CO INC.
PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR WPI; 97-225974/20.
PT Oligopeptide(s) recognised and cleaved by free prostate specific
PT antigen - useful in assays for active antigen, and in
PT oligopeptide-drug conjugates for prostatic cancer treatment
PS Disclosure; Fig 1; 192pp; English.
CC The present sequence represents Semenogelin I, a major sperm entrapping
CC gel protein. Oligopeptides designed based on the sequences surrounding
CC prostate specific antigen (PSA) cleavage sites (indicated above, and
CC numbered in order of the relative affinity of a site towards PSA
CC hydrolysis) are claimed. PSA, which has chymotrypsin-like specificity is
CC responsible for dissolution of the gel structure formed at ejaculation by
CC proteolysis of the major gel proteins (Semenogelin I and II, and
CC fibronectin). Liquefaction of the ejaculate releases progressively motile
CC spermatozoa. PSA complexed to alpha 1-antichymotrypsin is the predominant
CC form of serum PSA. Serum measurements of PSA are useful for monitoring
CC treatment of adenocarcinoma of the prostate. Prostate metastases are
CC also known to secrete immunologically reactive PSA since serum PSA is
CC detectable at high levels in prostatectomised patients showing widespread
CC metastatic prostate cancer. Cytotoxic compounds that could be activated
CC by the proteolytic activity of PSA should also be prostate cell specific
CC as well as specific for PSA secreting prostate metastases. Conjugates of
CC cytotoxic compounds and these oligopeptides are useful in treatment of
CC prostate cancer. The oligopeptides can also be used in a quantitative
CC assay for enzymatically active PSA.
SQ Sequence 462 AA;

Query Match 84.4%; Score 38; DB 26; Length 462;
Best Local Similarity 85.7%; Pred. No. 1.51e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 229 ehssklq 235
QY 1 EHSSKLQ 7

RESULT 5

ID R9583 standard; protein; 462 AA.
AC R9583;
DT 02-SEP-1996 (first entry)
DE Human semenogelin I
KW Human; semenogelin I; sperm entrapping gel; ejaculation; prostate; gel structure; dissolution; prostate specific antigen; proteolysis; chymotrypsin like specificity; peptide substrate; cleavage site; assay; determination; proteolytic activity; identification; inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 159..160
FT /note= "prostate specific antigen proteolytic cleavage site, CS#5"
FT cleavage_site 289..290
FT /note= "prostate specific antigen proteolytic cleavage site, CS#3"
FT cleavage_site 316..317
FT /note= "prostate specific antigen proteolytic cleavage site, CS#4"
FT cleavage_site 349..350
FT /note= "prostate specific antigen proteolytic cleavage site, CS#1"
FT cleavage_site 376..377
FT /note= "prostate specific antigen proteolytic cleavage site, CS#2"
FT WO9600503-Al.
PN 11-JAN-1996
PD 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI) MERCK & CO INC.
PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR WPI: 96-077275/08.
PT New peptide substrates cleaved by prostate-specific antigen - also cytototoxic conjugates for treating prostate cancer, and assay for determination of PSA activity
PT Example 1: Pages 50-51; 142pp; English.
PS The present sequence is human semenogelin I (hSI), one of the major proteins including hSI and fibronectin, in the sperm entrapping gel formed at ejaculation. This gel structure undergoes dissolution via the action of prostate specific antigen (PSA), a protease with chymotrypsin like specificity, which proteolyzes the above major proteins. New peptide substrates cleaved by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used in assays to determine the proteolytic activity of free PSA in a sample, and to identify cpts. which inhibit the proteolytic activity of PSA, they may also be conjugated, via a covalent bond or peptide linker, to a cytototoxic agent and used to treat prostate cancer. In a PSA hydrolysis assay, the percentage of hSI cleaved by YORK PSA after 30 mins. was 100 %.
CC Sequence 462 AA;
SQ

Query Match 84.4%; Score 38; DB 17; Length 462;
Best Local Similarity 85.7%; Pred. No. 1.51e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 229 ehskvq 235
:||||:
QY 1 EHSSKLQ 7

RESULT 6
ID W2182 standard; Protein; 432 AA.
AC W2182;
DT 13-JAN-1999 (first entry)
DE HSV-2 strain SB5 Contig ID 15 ORF15 protein.
KW HSV-2 strain SB5; immunological response induction; therapy; antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN WO9820016-Al.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
DR Esser KM, Leary JJ;
DR WPI: 98-286847/25.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention and treatment of infection or inducing immunological response in mammal
PS Claim 10: Page 50; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein sequence of the invention. This sequence was isolated from a HSV-2 strain SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 103.
CC Based on homology, this sequence is a 38K protein - human herpesvirus 1. The proteins can be used for the treatment or prevention of disease, to induce an immunological response in a mammal or to identify inhibitors, activators or novel antivirals. Antagonists of the proteins can be used to inhibit a viral polypeptide. The DNA sequence or a vector containing it can also be used to induce an immunological response in a mammal.
SQ Sequence 432 AA;
Query Match 80.0%; Score 36; DB 35; Length 497;
Best Local Similarity 42.9%; Pred. No. 2.80e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 166 dhtakle 172
:||||:
QY 1 EHSSKLQ 7

RESULT 7
ID W2030 standard; Protein; 497 AA.
AC W2030;
DT 07-DEC-1998 (first entry)
DE HSV-2 strain SB5 Contig ID 103 ORF8 protein.
KW HSV-2 strain SB5; immunological response induction; therapy; antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN WO9820016-Al.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
DR Esser KM, Leary JJ;
DR WPI: 98-286847/25.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention and treatment of infection or inducing immunological response in mammal
PS Claim 10: Page 50; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein sequence of the invention. This sequence was isolated from a HSV-2 strain SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 103.
CC Based on homology, this sequence is a 38K protein - human herpesvirus 1. The proteins can be used for the treatment or prevention of disease, to induce an immunological response in a mammal or to identify inhibitors, activators or novel antivirals. Antagonists of the proteins can be used to inhibit a viral polypeptide. The DNA sequence or a vector containing it can also be used to induce an immunological response in a mammal.
SQ Sequence 497 AA;
Query Match 80.0%; Score 36; DB 35; Length 497;
Best Local Similarity 42.9%; Pred. No. 2.80e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 dhtakle 107
:||||:
QY 1 EHSSKLQ 7

PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
DR Esser KM, Leary JJ;
DR WPI: 98-286847/25.
DR N-PSDB; V62176.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention and treatment of infection or inducing immunological response in mammal
PS Claim 10: Page 115; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein sequence of the invention. This sequence was isolated from a HSV-2 strain SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
CC Based on homology, this sequence is a probable DNA packaging protein. The proteins can be used for the treatment or prevention of disease, to induce an immunological response in a mammal or to identify inhibitors, activators or novel antivirals. Antagonists of the proteins can be used to inhibit a viral polypeptide. The DNA sequence or a vector containing it can also be used to induce an immunological response in a mammal.
SQ Sequence 432 AA;

Query Match 80.0%; Score 36; DB 36; Length 432;
Best Local Similarity 42.9%; Pred. No. 2.80e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 dhtakle 107
:||||:
QY 1 EHSSKLQ 7

RESULT 7
ID W2030 standard; Protein; 497 AA.
AC W2030;
DT 07-DEC-1998 (first entry)
DE HSV-2 strain SB5 Contig ID 103 ORF8 protein.
KW HSV-2 strain SB5; immunological response induction; therapy; antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN WO9820016-Al.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
DR Esser KM, Leary JJ;
DR WPI: 98-286847/25.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention and treatment of infection or inducing immunological response in mammal
PS Claim 10: Page 50; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein sequence of the invention. This sequence was isolated from a HSV-2 strain SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 103.
CC Based on homology, this sequence is a 38K protein - human herpesvirus 1. The proteins can be used for the treatment or prevention of disease, to induce an immunological response in a mammal or to identify inhibitors, activators or novel antivirals. Antagonists of the proteins can be used to inhibit a viral polypeptide. The DNA sequence or a vector containing it can also be used to induce an immunological response in a mammal.
SQ Sequence 497 AA;

Query Match 80.0%; Score 36; DB 35; Length 497;
Best Local Similarity 42.9%; Pred. No. 2.80e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 166 dhtakle 172
:||||:
QY 1 EHSSKLQ 7

```

CC or homologue to come into association with a test compound; (C)
CC providing UL15 of HSV-1 or a functional derivative or homologue
CC thereof; (d) screening for inhibition of the interaction between the
CC VP5 or its derivative or homologue and UL15 or its derivative or
CC homologue, where such inhibition is indicative that the test compound
CC is an antiviral agent. The present sequence represents the HSV-1 UL15
CC protein.
SQ Sequence 735 AA;

Query Match 80.0%; Score 36; DB 34; Length 735;
Best Local Similarity 42.9%; Pred. No. 2.80e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 dhtakle 107
QY 1 EHSSKLQ 7

RESULT 10
ID R08121 standard; protein; 212 AA.
AC R08121;
DT 22-JAN-1991 (first entry)
DE Modified tilapia prolactin I.
KW Fish; hydromineral balance; pituitary; hormone; osmoregulation;
TPRLI.
OS Oreochromis niloticus.
FH Key Location/Qualifiers
FT peptide 1..24
FT FT /label=signal peptide
FT protein 25..212
FT FT /label=TPRLI
FT FT /note="calculated as 20.9 kD"
PN EP-387457-A.
PD 19-SEP-1990.
PF 06-JAN-1989; 400047.
PR 06-JAN-1989; EP-400047.
PA (EURO-) EUROGENTEC SA.
PI RentierDelrue F, Martial J, Renard A;
DR WPI: 90-283784/38.
DR N-PSDB: Q06742.
PT New recombinant fish polypeptide(s) - having growth and/or
PT osmoregulatory properties for enhancing growth and adaption of
PT fish cultures
PT Claim 10; Fig 2bis; 29pp; English.
CC The sequence was deduced from a cDNA clone isolated from a library
CC prep'd. from mRNA extracted from the anterior pituitary gland of
CC tilapia. It encodes a modified prolactin (TPRLI) which is thought
CC to play an important role in osmoregulation (hydromineral balance)
CC in this species. The gene can be expressed using the strong
CC promoter of gene 10 of bacteriophage T7. The gene was modified to
CC remove the signal peptide coding region and the vector expressed
CC in E.coli HB101. The protein differs from R06893 at only two
CC positions: Leu and Ser replace Phe and Thr at AAs 7 and 100 resp.
CC The product can be admin. to fish cultures to enhance growth and
CC improve their adaptation to culturing conditions. Dosage is 5
CC ng-1 ug protein/g fish/day by injection or 1 ug-1 mg protein/g
CC fish food/day.
CC See also R06893, R06914 and R08122.
SQ Sequence 212 AA;

Query Match 77.8%; Score 35; DB 2; Length 212;
Best Local Similarity 57.1%; Pred. No. 3.81e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 36 qhsdklh 42
QY 1 EHSSKLQ 7

RESULT 11
ID R06893 standard; protein; 212 AA.
AC R06893;
DT 22-JAN-1991 (first entry)

```

DE Tilapia prolactin I.
KW Fish; hydromineral balance; pituitary; hormone; osmoregulation;
OS Oreochromis niloticus.
FH Key Location/Qualifiers
FT peptide 1..24
FT protein /label=signal peptide
FT /label=TPRLI
FT /note="calculated as 20.9 kD"
PN EP-387457-A.
PD 19-SEP-1990.
PF 06-JAN-1989; 400047.
PR 06-JAN-1989; EP-400047.
PA (EURO-) EUROGENTEC SA.
PI Rentier-Delrue F, Martial J, Renard A;
DR WPI: 90-283784/39.
DR N-PSDB: Q05981.
PT New recombinant fish polypeptide(s) - having growth and/or
PT osmoregulatory properties for enhancing growth and adaptation of
PT fish cultures
PS Claim 10; Fig 2; 29pp; English.
CC The sequence was deduced from a cDNA clone isolated from a library
CC prep'd. from mRNA extracted from the anterior pituitary gland of
CC tilapia. It encodes a prolactin (TPRLI) which is thought to play
CC an important role in osmoregulation (hydromineral balance) in this
CC species. The gene can be expressed using the strong promoter of
CC gene 10 of bacteriophage T7. The gene was modified to remove the
CC signal peptide coding region and the vector expressed in E.coli
CC HB101. The product can be admin. to fish cultures to enhance
CC growth and improve their adaptation to culturing conditions.
CC Dosage is 5 ng-1 ug protein/g fish/day by injection or 1 ug-1 mg
CC protein/g fish food/day.
CC See also R06914, R08121 and R08122.
SQ Sequence 212 AA;

Query Match 77.8%; Score 35; DB 2; Length 212;
Best Local Similarity 57.1%; Pred. No. 3.81e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 36 qhsdklh 42
:|||
Qy 1 EHSSKLQ 7

RESULT 12
ID R74094 standard; Protein; 742 AA.
AC R74094;
DT 04-JAN-1996 (first entry)
DE Human zona pellucida-2 (H2P-2) protein.
KW Human zona pellucida; H2P-2; protein; vaccine; antigen;
KW contraceptive; ovary.
OS Homo sapiens.
PN J07099974-A.
PD 18-APR-1995.
PF 05-OCT-1993; 249404.
PR 05-OCT-1993; JP-249404.
PA (TOFU) TONEN CORP.
DR WPI: 95-182067/24.
DR N-PSDB: Q92234.
PT A DNA coding human zona pellucida-2 protein - used as a vaccine
PT antigen
PS Claim 1; Page 8-11; 11pp; Japanese.
CC The human zona pellucida-2 (H2P-2) protein may be used as a vaccine
CC antigen. It can be artificially synthesised using recombinant
CC techniques. Partial sequences of the H2P-2 gene were subcloned, the
CC sequences and primers used in cloning and subcloning are given in
CC Q92255-69.
SQ Sequence 742 AA;

Query Match 77.8%; Score 35; DB 14; Length 742;
Best Local Similarity 56.7%; Pred. No. 3.81e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 472 hatklq 477
:|||
Qy 2 HSSKLQ 7

RESULT 13
ID R25138 standard; Protein; 2431 AA.
AC R25138;
DT 05-JAN-1993 (first entry)
DE SFV4 non-structural polyprotein.
KW Semliki forest virus; SFV; SP6 expression vector; RNA polymerase.
OS Synthetic.
PN WO9210578-A.
PD 25-JUN-1992.
PF 12-DEC-1991; SE0855.
PR 13-DEC-1990; SE-003978.
PA (BIOP-) BIOTION AB.
PI Garoff H, Liljestrom P;
DR WPI: 92-234633/28.
DR N-PSDB: Q26021.
PT RNA mol. derived from alphavirus RNA genome - chimeric alphavirus
PT antigen and vaccine for immunisation against viral infections
PS Disclosure; Fig 5; 94pp; English.
CC The sequence given is the non-structural polyprotein encoded by a full
CC length Semliki forest virus (SFV) cDNA clone within an SP6 expression
CC vector. The SP6 RNA polymerase promoter allows in vitro transcription
CC of full length and infectious transcripts.
SQ Sequence 2431 AA;

Query Match 77.8%; Score 35; DB 5; Length 2431;
Best Local Similarity 57.1%; Pred. No. 3.81e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1237 dhamklq 1243
:|||
Qy 1 EHSSKLQ 7

RESULT 14
ID W50891 standard; Protein; 3084 AA.
AC W50891;
DT 07-DEC-1998 (first entry)
DE Mouse laminin A chain.
KW Laminin; mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
KW malignancy; Familial Mediterranean Fever; multiple myeloma;
KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
KW Gerstmann-Straussler syndrome; kuru; scrapie; haemodialysis;
KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
KW therapy.
OS Mus sp.
PN J07099974-A.
PD 18-APR-1995.
PF 05-OCT-1993; 249404.
PR 05-OCT-1993; JP-249404.
PA (TOFU) TONEN CORP.
DR WPI: 95-182067/24.
DR N-PSDB: Q92234.
PT A DNA coding human zona pellucida-2 protein - used as a vaccine
PT antigen
PS Claim 1; Page 8-11; 11pp; Japanese.
CC The human zona pellucida-2 (H2P-2) protein may be used as a vaccine
CC antigen. It can be artificially synthesised using recombinant
CC techniques. Partial sequences of the H2P-2 gene were subcloned, the
CC sequences and primers used in cloning and subcloning are given in
CC Q92255-69.
SQ Sequence 742 AA;

Query Match 77.8%; Score 35; DB 14; Length 742;
Best Local Similarity 56.7%; Pred. No. 3.81e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

CC inhibitors of amyloid formation, deposition, accumulation and/or
 CC persistence in Alzheimer's disease and other amyloidoses. The
 CC laminin products (see W50888-98) may include mouse or human laminin
 CC A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merosin),
 CC laminin G1 chain, the globular repeats of the laminin A1 chain and
 CC the beta-amyloid binding domain of the laminin A chain. A claimed
 CC method for treating an amyloid disease comprises administering a
 CC polypeptide having a conformational similarity to a fragment of a
 CC laminin protein. A method for diagnosing an amyloid disease
 CC involves determining levels of laminin in a sample. Production
 CC of laminin or its fourth globular repeat in vivo provides a method
 CC for in vivo inhibition of beta-amyloid amyloidosis. The products
 CC and methods can be used for the diagnosis, prognosis, monitoring
 CC and treatment of amyloidoses such as Alzheimer's disease, Down's
 CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
 CC the Dutch type (where the specific amyloid is the beta-amyloid
 CC protein), the amyloidosis associated with chronic inflammation,
 CC various forms of malignancy and Familial Mediterranean Fever (AA
 CC amyloid or inflammation-association amyloidosis); the amyloidosis
 CC associated with multiple myeloma and other B-cell abnormalities
 CC (AL amyloid), the amyloidosis associated with type II diabetes
 CC (amylin or islet amyloid), the amyloidosis associated with prion
 CC diseases including Creutzfeldt-Jacob disease, Gertsmann-Straussler
 CC syndrome, kuru and animal scrapie (Prp amyloid), the amyloidosis
 CC associated with long-term haemodialysis and carpal tunnel syndrome
 CC (beta 2-microglobulin amyloid), the amyloidosis associated with
 CC senile cardiac amyloid and Familial Amyloidotic Polyneuropathy
 CC (prealbumin or transthyretin amyloid), and the amyloidosis
 CC associated with endocrine tumours such as medullary carcinoma of
 CC the thyroid (variant of procalcitonin).
 CC Sequence 3084 AA;

Query Match 77.8%; Score 35; DB 35; Length 3084;
 Best Local Similarity 71.4%; Pred. No. 3.81e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1763 nhseklq 1769
 :|||
 QY 1 EHSSKLQ 7

RESULT 15
 ID W27419 standard; peptide; 7 AA.
 AC W27419;
 DT 19-DEC-1997 (first entry)
 DE CDR2 from light chain variable region of KM1259 antibody.
 KW Complementarity determining region; CDR; light chain; treatment;
 KW variable region; murine; mouse; human; interleukin 5; IL-5;
 KW receptor; alpha chain; monoclonal antibody; hybridoma; detection;
 KW assay; diagnosis; allergic respiratory disease; chronic bronchitis.
 OS Mus spp.
 PN WO9710354-A1.
 PD 20-MAR-1997.
 PF 11-SEP-1996; J02588.
 PR 11-SEP-1995; JP-232384.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Anazawa H, Furuya A, Hanai N, Iida A, Koike M;
 PI Nakamura K, Takatsu K;
 DR WPI; 97-202249/18.
 PT Antibody against alpha-chain of human interleukin 5 receptor -
 PT useful for diagnosis and treatment of respiratory allergic diseases,
 PT e.g. chronic bronchitis
 PS Claim 8; Page 165; 238pp; Japanese.
 CC The present sequence is complementarity determining region 2 (CDR2)
 CC from the light chain variable region of the murine anti-human
 CC interleukin 5 receptor alpha chain (hIL-5R alpha) monoclonal
 CC antibody (Mab) KM1259. KM1259 is produced by the hybridoma
 CC FERM BP-5134, which was prepared by immunising Balb/c mice with
 CC hIL-5R alpha, fusing spleen cells obtained from the mice with mouse
 CC myeloma P3-U1 cells and screening the resultant hybridomas. The Mab
 CC can be used to detect or assay for hIL-5R alpha and cells
 CC expressing it on their surface, especially to diagnose allergic
 CC respiratory diseases, e.g. chronic bronchitis. It can also be used

CC to treat such diseases.
 SQ Sequence 7 AA;

Query Match 75.6%; Score 34; DB 24; Length 7;
 Best Local Similarity 66.7%; Pred. No. 5.15e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 htsrlq 6
 :|:|
 QY 2 HSSKLQ 7

Search completed: Thu Oct 28 11:24:36 1999
 Job time : 20 secs.

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:24:52 1999; Maspar time 2.96 Seconds

Tabular output not generated. 94.825 Million cell updates/sec

Title: >US-09-081-707-10
Description: (1-7) from US09081707.pap
Perfect Score: 45
Sequence: 1 EHSSKLQ 7

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r60
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 18.827; Variance 17.198; scale 1.095

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	100.0	582	2 A43412	semenogelin II precursor	1.66e+01
2	42	93.3	509	2 T01344	hypothetical protein	1.20e+00
3	42	93.3	675	3 T00013	DAP-1 beta protein -	1.20e+00
4	42	93.3	692	2 T00025	PSD-95 binding protein	1.20e+00
5	42	93.3	977	3 T00014	DAP-1 alpha protein -	1.20e+00
6	40	88.9	265	2 S63532	NAD(P)H-quinone oxidoreductase	4.25e+00
7	40	88.9	585	2 G64220	ATP-binding protein m	4.25e+00
8	39	86.7	147	2 S13157	hemoglobin - polychaete	7.87e+00
9	39	86.7	585	2 E64892	probable membrane protein	7.87e+00
10	38	84.4	210	2 A49213	major surface protein	1.44e+01
11	38	84.4	224	2 G71281	probable Mg2+ transpore	1.44e+01
12	38	84.4	337	2 B44478	probable cell growth	1.44e+01
13	38	84.4	373	2 A44478	probable cell growth	1.44e+01
14	38	84.4	420	2 A54759	cytochrome ba(3) chain	1.44e+01
15	38	84.4	462	1 WTH0B	semenogelin I precursor	1.44e+01
16	38	84.4	556	2 T03114	tegument protein - al	1.44e+01
17	38	84.4	617	2 T02121	hypothetical protein	1.44e+01
18	38	84.4	664	2 S58162	hypothetical protein	1.44e+01
19	37	82.2	91	2 E5011	hypothetical protein	2.61e+01
20	37	82.2	278	2 H55089	hypothetical protein	2.61e+01
21	37	82.2	356	1 S55437	translocation releasing	2.61e+01
22	37	82.2	604	2 S66993	hypothetical protein	2.61e+01
23	37	82.2	700	2 A61527	stonustoxin beta chain	2.61e+01

24	37	82.2	883	2 A71434	probable RNA helicase	2.61e+01
25	37	82.2	2479	1 MNWVRA	nonstructural polypro	2.61e+01
26	36	80.0	147	2 B36529	hemoglobin P3 - polyc	4.66e+01
27	36	80.0	265	2 S22297	gene DdLRep3 protein	4.66e+01
28	36	80.0	301	2 E71855	ATP synthase F1, chain	4.66e+01
29	36	80.0	301	2 E64661	H+-transporting ATP s	4.66e+01
30	36	80.0	324	2 S28672	occr protein - Agrobac	4.66e+01
31	36	80.0	340	2 G70741	hypothetical protein	4.66e+01
32	36	80.0	344	1 WMBE31	38k protein - human h	4.66e+01
33	36	80.0	491	2 B70414	NADH dehydrogenase I	4.66e+01
34	36	80.0	531	2 B54096	flavin-containing mon	4.66e+01
35	36	80.0	533	1 S18380	dimethylalanine monoo	4.66e+01
36	36	80.0	735	1 WMBE5	UL15 protein - human	4.66e+01
37	36	80.0	744	2 S45060	outer capsid spike pr	4.66e+01
38	36	80.0	744	2 S45061	outer capsid spike pr	4.66e+01
39	36	80.0	764	2 I51302	myosin heavy chain -	4.66e+01
40	36	80.0	875	2 S70115	ZIP1 protein - yeast	4.66e+01
41	36	80.0	980	2 E71606	hypothetical protein	4.66e+01
42	36	80.0	1381	2 S45781	probable calcium-bind	4.66e+01
43	36	80.0	4466	1 S17231	dynein beta heavy cha	4.66e+01
44	35	77.8	948	2 A57640	retinoblastoma protei	8.21e+01
45	35	77.8	1493	2 F70435	glutamate synthase la	8.21e+01

ALIGNMENTS

RESULT 1

ENTRY	A43412	#type complete
TITLE	semenogelin II precursor - human	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-May-1998	
ACCESSIONS	A43412; B31489; A45295; S29156; S68765; S68762	
REFERENCE	A43412; B31489; A45295; S29156; S68765; S68762	
#authors	Ulvback, M.; Lazure, C.; Lilja, H.; Spurr, N.K.; Rao, V.V.; Loffler, C.; Hansmann, I.; Lundwall, A.	
#journal	J. Biol. Chem. (1992) 267:18080-18084	
#title	Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20.	
#cross-references	MUID:92388176	
#accession	A43412	
#molecule_type	DNA	
#residues	1-582 #label ULV	
#cross-references	GB:M81651; NID:g307417; PID:g307418	
#note	Sequence extracted from NCBI backbone (NCBI:112887, NCBI:112889)	
REFERENCE	A31489	
#authors	Lilja, H.; Abrahamsson, P.A.; Lundwall, A.	
#journal	J. Biol. Chem. (1989) 264:1894-1900	
#title	Seamenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa.	
#cross-references	MUID:89109215	
#accession	B31489	
#status	nucleic acid sequence not shown	
#molecule_type	mRNA	
#residues	214-278, 'Y', 280-281 #label LIL	
REFERENCE	A45295	
#authors	Lilja, H.; Lundwall, A.	
#journal	Proc. Natl. Acad. Sci. U.S.A. (1992) 89:4559-4563	
#title	Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein.	
#cross-references	MUID:92262479	
#accession	A45295	
#molecule_type	mRNA	
#residues	3-582 #label LI2	
#cross-references	GB:M81652	
REFERENCE	S29155	
#authors	Schneider, K.; Kausler, W.; Triptier, D.; Jouvenal, K.; Spittler, G.	
#journal	Biol. Chem. Hoppe-Seyler (1989) 370:353-356	
#title	Isolation and structure determination of two peptides	

```

#accession      occurring in human seminal plasma.
#molecule_type protein
#residues      390,'E',392-396,'E',398-400,'EW',403-405,'C',407,'DE'
#label SCH
#note          this report is of a secondary sequence determined
                simultaneously with the sequence with accession number
                S29155 (see entry WTHUB); as a secondary sequence, it
                should be considered less than fully reliable

S68765
#authors      Kise, H.; Nishioka, J.; Kawamura, J.; Suzuki, K.
#journal      Eur. J. Biochem. (1996) 238:88-96
#title        Characterization of semenogelin II and its molecular
                interaction with prostate-specific antigen and protein C
                inhibitor.
#cross-references MUID:96248425
#accession    S68765
#molecule_type protein
#residues     105-107,'K',109-111,113-122;260-269;280-283 #label KIS
REFERENCE     S68761
#authors      Malm, J.; Hellman, J.; Magnusson, H.; Laurell, C.B.; Lilja,
                H.
#journal      Eur. J. Biochem. (1996) 238:48-53
#title        Isolation and characterization of the major gel proteins in
                human semen, semenogelin I and semenogelin II.
#accession    S68762
#molecule_type protein
#residues     420-421,'G',423-423 #label MAL
GENETICS
#gene         GDB:SEMG2
#cross-references GDB:132657; OMIM:182141
#map_position 20q12-20q13.1
#introns      26/1
CLASSIFICATION #superfamily semenogelin
KEYWORDS       duplication; glycoprotein; semen; seminal vesicle; tandem
                repeat
FEATURE
1-23          #domain signal sequence #status predicted #label SIG\
24-582        #product semenogelin II #status predicted #label MAT\
272           #binding_site carbohydrate (Asn) (covalent) #status
                predicted
SUMMARY        #length 582 #molecular-weight 65444 #checksum 9704
Query Match    100.0%; Score 45; DB 2; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.66e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 229 EHSSKQ 235
QY 1 EHSSKQ 7

RESULT 2
ENTRY      T01344 #type complete
TITLE      hypothetical protein F6N15.5 - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
                cress
DATE       12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                12-Feb-1999
ACCESSIONS T01344
REFERENCE   Z14297
#authors    Ryan, E.; Edwards, J.; Pape, K.
#submission Submitted to the EMBL Data Library, May 1998
#description The sequence of A. thaliana F6N15.
#accession  T01344
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues   1-509 #label RYA
#cross-references EMBL:AF069299; NID:g3193311; PID:g3193321
GENETICS
#map_position 50/2; 342/3; 399/3; 496/2
#introns      F6N15.5
#note

SUMMARY        #length 509 #molecular-weight 58562 #checksum 4630
Query Match    93.3%; Score 42; DB 2; Length 509;
Best Local Similarity 85.7%; Pred. No. 1.20e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 375 EHSPKQ 381
QY 1 EHSSKQ 7

RESULT 3
ENTRY      T00013 #type complete
TITLE      DAP-1 beta protein - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       19-Jan-1999 #sequence_revision 19-Jan-1999 #text_change
                19-Jan-1999
ACCESSIONS T00013
REFERENCE   Z14044
#authors    Satoh, K.; Yanai, H.; Senda, T.; Kohu, K.; Nakamura, T.;
                Okumura, N.; Matsumine, A.; Kobayashi, S.; Toyoshima, K.;
                Akiyama, T.
#journal      Genes to Cells (1997) 2:415-424
#title        DAP-1, a novel protein that interacts with the guanylate
                kinase-like domains of hDLG and PSD-95.
#accession    T00013
#status       preliminary
#residues     1-675 #label SAT
#cross-references EMBL:AB000276; NID:d1173041; PID:d1024134
SUMMARY        #length 675 #molecular-weight 75260 #checksum 1624
Query Match    93.3%; Score 42; DB 3; Length 675;
Best Local Similarity 85.7%; Pred. No. 1.20e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 93 EHSPKQ 99
QY 1 EHSSKQ 7

RESULT 4
ENTRY      T00025 #type complete
TITLE      PSD-95 binding protein - rat
ORGANISM   #formal_name Rattus norvegicus #common_name Norway rat
DATE       22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                22-Jan-1999
ACCESSIONS T00025
REFERENCE   Z14063
#authors    Kawashima, N.; Takamiya, K.; Sun, J.; Kitabatake, A.; Sobue,
                K.
#journal      FEBS Lett. (1997) 418:301-304
#title        Differential expression of isoforms of PSD-95 binding protein
                (GKAP/SAPAP1) during rat brain development.
#accession    T00025
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-692 #label KAW
#cross-references EMBL:AB003594; NID:d1177606; PID:d1025176
SUMMARY        #length 692 #molecular-weight 76991 #checksum 9411
Query Match    93.3%; Score 42; DB 2; Length 692;
Best Local Similarity 85.7%; Pred. No. 1.20e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 101 EHSPKQ 107
QY 1 EHSSKQ 7

RESULT 5
ENTRY      T00014 #type complete
TITLE      DAP-1 alpha protein - human
ORGANISM   #formal_name Homo sapiens #common_name man
```

```

DATE      19-Jan-1999 #sequence_revision 19-Jan-1999 #text_change
ACCESSIONS T00014
REFERENCE   Z14044
#authors   Sato, K.; Yanai, H.; Senda, T.; Kohu, K.; Nakamura, T.;
           Okumura, N.; Matsumine, A.; Kobayashi, S.; Toyoshima, K.;
           Akiyama, T.
#journal   Genes to Cells (1997) 2:415-424
#title     DAP-1, a novel protein that interacts with the guanylate
           kinase-like domains of HDLG and PSD-95.
#accession T00014
#status    preliminary
#residues  1-977 ##label SAT
##cross-references EMBL:AB000277; NID:d1173042; PID:d1024135
SUMMARY    #length 977 #molecular-weight 108872 #checksum 4168

Query Match      93.3%; Score 42; DB 3; Length 977;
Best Local Similarity 85.7%; Pred. No. 1.20e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 395 EHSKQLQ 401
   |||:||||
QY 1 EHSSKQLQ 7

RESULT      6
ENTRY       S63532
TITLE       NAD(P)H-quinone oxidoreductase, 23K, precursor - Narcissus
           pseudonarcissus
ORGANISM    #formal_name Narcissus pseudonarcissus
DATE        19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
ACCESSIONS  S63532; S63488; S54136
REFERENCE    S63488
#authors     Niveststein, V.; Vandekerckhove, J.; Tadros, M.H.; Lintig,
           J.V.; Nitschke, W.; Beyer, P.
#journal     Eur. J. Biochem. (1995) 233:864-872
#title       Carotene desaturation is linked to a respiratory redox
           pathway in Narcissus pseudonarcissus chromoplast membranes:
           involvement of a 23-kDa oxygen-evolving-complex-like
           protein.
##cross-references MUID:96085151
#accession  S63532
#molecule_type mRNA
#residues   1-265 ##label NIE
##cross-references EMBL:X78816; NID:g780272; PID:g780273
##experimental_source Chromoplast
#accession  S63488
#molecule_type protein
#residues   80-89;93-106;128-148;170-175;247-249 ##label NIW

GENETICS
#genome     nuclear
CLASSIFICATION #superfamily photosystem II oxygen-evolving complex protein 2
KEYWORDS     thylakoid
FEATURE      1-79
           #domain transit peptide (chromoplast) #status
           experimental #label TNP\
80-265      #product NAD(P)H-quinone oxidoreductase, 23K #status
           experimental #label MAT
SUMMARY      #length 265 #molecular-weight 28521 #checksum 5299

Query Match      88.9%; Score 40; DB 2; Length 265;
Best Local Similarity 71.4%; Pred. No. 4.25e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 35 EHSARLQ 41
   |||:||||
QY 1 EHSSKQLQ 7

RESULT      7
ENTRY       G64220
TITLE       ATP-binding protein msMK homolog - Mycoplasma genitalium

```

```

(SGC3)
#formal_name Mycoplasma genitalium
17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
17-Jul-1998
ACCESSIONS G64220
REFERENCE   A64200
#authors    Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
           R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
           Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
           Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
           Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
           J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
           Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
           C.A.; Venter, J.C.
#journal     Science (1995) 270:397-403
#title       The minimal gene complement of Mycoplasma genitalium.
##cross-references MUID:96026346
#accession  G64220
#status     preliminary; nucleic acid sequence not shown;
           translation not shown
##molecule_type DNA
#residues    1-585 ##label TIGR
##cross-references GB:U39696; GB:U43967; NID:gl045869; PID:gl045872;
           TIGR:MG187
##experimental_source strain G-37

GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily ATP-binding cassette homology
KEYWORDS      ATP; P-loop
FEATURE       23-444
           #domain ATP-binding cassette homology #status atypical
           #label ABCL\
40-47        #region nucleotide-binding motif A (P-loop)
SUMMARY      #length 585 #molecular-weight 66948 #checksum 9140

Query Match      88.9%; Score 40; DB 2; Length 585;
Best Local Similarity 71.4%; Pred. No. 4.25e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 225 EHNAKLQ 231
   |||:||||
QY 1 EHSSKQLQ 7

RESULT      8
ENTRY       S13157
TITLE       hemoglobin - polychaete (Glycera dibranchiata)
ORGANISM    #formal_name Glycera dibranchiata #common_name bloodworm
DATE        19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
           26-Feb-1998
ACCESSIONS  S13157
REFERENCE    S13157
#authors     Zafar, R.S.; Chow, L.H.; Stern, M.S.; Vinogradov, S.N.; Walz,
           D.A.
#journal     Biochim. Biophys. Acta (1990) 1041:117-122
#title       The heterogeneity of the polymeric intracellular hemoglobin
           of Glycera dibranchiata and the cDNA-derived amino acid
           sequence of one component.
##cross-references MUID:91091412
#accession  S13157
#status     preliminary
#molecule_type mRNA
#residues    1-147 ##label ZAF
CLASSIFICATION #superfamily globin; globin homology
SUMMARY      #length 147 #molecular-weight 16029 #checksum 7221

Query Match      86.7%; Score 39; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 7.87e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 57 HSSKQLQ 62
   |||:||||
QY 2 HSSKQLQ 7

```



```

(alternatively spliced type I transcript) - human
#formal_name Homo sapiens #common_name man
10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
A44478
A44478
Weitzel, J.N.; Kasperczyk, A.; Mohan, C.; Krontiris, T.G.
Genomics (1992) 14:309-319
The HRA51 gene cluster: two upstream regions recognizing
transcripts and a third encoding a gene with a leucine
zipper domain.
#cross-references MUID:93052330
#accession A44478
##status preliminary; not compared with conceptual translation
##molecule_type nucleic acid
##residues 1-373 ##label WEI
##cross-references GB:M91083; NID:gl84389; PID:gl84390
##note sequence extracted from NCBI backbone (NCBIP:117869)
SUMMARY
length 373 #molecular_weight 39945 #checksum 3379
Query Match 84.4%; Score 38; DB 2; Length 373;
Best Local Similarity 57.1%; Pred. NO. 1.44e-01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 202 EHAARLQ 208
QY 1 EHSSKQ 7
II:::II

RESULT 14
ENTRY A54759 #type complete
TITLE cytochrome ba(3) chain II precursor - Paracoccus
denitrificans
ORGANISM #formal_name Paracoccus denitrificans
DATE 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change
09-Sep-1997
A54759
A54759
Richter, O.M.H.; Tao, J.; Turba, A.; Ludwig, B.
J. Biol. Chem. (1994) 269:23079-23086
A cytochrome ba-3 functions as a quinol oxidase in Paracoccus
denitrificans. Purification, cloning, and sequence
comparison.
#accession A54759
##status preliminary
##molecule_type DNA
##residues 1-420 ##label RIC
##cross-references GB:X78196; NID:g461327; PID:g673470
##note it is uncertain whether Met-1 or Met-21 is the
initiator; Met-21 seems more likely
GENETICS
#gene qoxA
SUMMARY #length 420 #molecular_weight 46529 #checksum 4989
Query Match 84.4%; Score 38; DB 2; Length 420;
Best Local Similarity 57.1%; Pred. NO. 1.44e-01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 100 DHSTKLE 106
QY 1 EHSSKQ 7
:::III:

RESULT 15
ENTRY WTHUB #type complete
TITLE semenogelin I precursor - human
CONTAINS seminal basic protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 04-Dec-1986 #sequence_revision 03-Oct-1995 #text_change
26-Feb-1999
B43412; A31489; A91335; S29155; A43500; A91320; S29380;
S68761; A03254
A43412
REFERENCE

```

```

#authors Ulvback, M.; Lazure, C.; Lilja, H.; Spurr, N.K.; Rao, V.V.;
Loffler, C.; Hansmann, I.; Lundwall, A.
#journal J. Biol. Chem. (1992) 267:18080-18084
#title Gene structure of semenogelin I and II. The predominant
proteins in human semen are encoded by two homologous genes
on chromosome 20.
#cross-references MUID:92388176
#accession B43412
##molecule_type DNA
##residues 1-462 ##label ULV
##cross-references GB:M81650; NID:g307416; PID:g487420
REFERENCE A31489
#authors Lilja, H.; Abrahamsson, P.A.; Lundwall, A.
#journal J. Biol. Chem. (1989) 264:1894-1900
#title Semenogelin, the predominant protein in human semen. Primary
structure and identification of closely related proteins in
the male accessory sex glands and on the spermatozoa.
#cross-references MUID:89109215
#accession A31489
##molecule_type mRNA
##residues 1-78, 'T', 80-422, 'K', 424-462 ##label LIL
##cross-references GB:J04440
##note Ser-79 was also found
REFERENCE A91335
#authors Lilja, H.; Jeppsson, J.O.
#journal FEBS Lett. (1985) 182:181-184
#title Amino acid sequence of the predominant basic protein in human
seminal plasma.
#cross-references MUID:85127550
#accession A91335
##molecule_type protein
##residues 108-159 ##label LI2
##note this sequence represents a naturally occurring fragment
from proteolytic cleavage of semenogelin during
liquefaction of semen
REFERENCE S29155
#authors Schneider, K.; Kausler, W.; Tripiet, D.; Jouvénal, K.;
Spiteller, G.
#journal Biol. Chem. Hoppe-Seyler (1989) 370:353-356
#title Isolation and structure determination of two peptides
occurring in human seminal plasma.
#accession S29155
##molecule_type protein
##residues 316-320, 'L', 322-344 ##label SCH
##note this sequence represents the amino end of a naturally
occurring fragment from proteolytic cleavage of
semenogelin during liquefaction of semen
REFERENCE A43500
#authors Ramasharma, K.; Sairam, M.R.; Seidah, N.G.; Chretien, M.;
Manjunath, P.; Schiller, P.W.; Yamashiro, D.; Li, C.H.
#journal Science (1984) 223:1199-1202
#title Isolation, structure, and synthesis of a human seminal plasma
peptide with inhibin-like activity.
#cross-references MUID:84146751
#accession A43500
##molecule_type protein
##residues 108-138 ##label RAM
##note this sequence represents a naturally occurring fragment
from proteolytic cleavage of semenogelin during
liquefaction of semen
REFERENCE A91320
#authors Seidah, N.G.; Ramasharma, K.; Sairam, M.R.; Chretien, M.
#journal FEBS Lett. (1984) 167:98-102
#title Partial amino acid sequence of a human seminal plasma peptide
with inhibin-like activity.
#cross-references MUID:84132557
#accession A91320
##molecule_type protein
##residues 108-138 ##label SEI
##note this sequence represents a naturally occurring fragment
from proteolytic cleavage of semenogelin during
liquefaction of semen
REFERENCE S29380

```

```
#authors      Khan, Z.; Smyth, D.G.
#journal      Eur. J. Biochem. (1993) 212:35-40
#title        Isolation and identification of N-terminally extended forms
              of 5-oxoprollylglutamylprolinamide (Glp-Glu-Pro-NH(2)), a
              thyrotropin-releasing-hormone(TRH)-like peptide present in
              human semen.
#cross-references MUID:93185635
#accession     S29380
              ##molecule_type protein
              ##residues      373-397 #label KHA
              ##note          the authors' suggestion that this peptide is amidated is
                              consistent with radioimmunoassay results but has not
                              been proven
              ##note          the amidated tripeptide Glp-Glu-Pro-NH2 (where Glp is
                              pyroglutamic acid) is present in human semen; although
                              the sequence reported here is consistent with an
                              amino-terminally extended form derived from
                              semenogelin, the sequence is followed in semenogelin
                              by Trp, which (unlike Gly) probably cannot become the
                              source of an amide moiety; the authors conclude the
                              peptide must be derived from a closely related protein

REFERENCE      S68761
#authors      Malm, J.; Hellman, J.; Magnusson, H.; Laurell, C.B.; Lilja,
              H.
#journal      Eur. J. Biochem. (1996) 238:48-53
#title        Isolation and characterization of the major gel proteins in
              human semen, semenogelin I and semenogelin II.
#accession     S68761
              ##molecule_type protein
              ##residues      49-50,'G',52-53 #label MAL
COMMENT        This abundant protein from seminal vesicle secretions maintains a
              gel-like environment for the sperm cells. At ejaculation,
              kallikrein-like enzymes in prostatic secretions cleave this
              protein, resulting in liquefaction of the seminal gel and
              allowing increased sperm motility.

GENETICS
#gene          GDB:SPMG1
#cross-references GDB:128167; OMIM:182140
#map_position   20q12-20q13.1
#introns        26/1
CLASSIFICATION #superfamily semenogelin
KEYWORDS        duplication; glycoprotein; semen; seminal vesicle
FEATURE
1-24            #domain signal sequence #status predicted #label SIG\
108-138         #product seminal basic protein #status experimental
              #label MAT\
174-215         #region semenogelin short repeat 1\
234-275         #region semenogelin short repeat 2\
282-339         #region semenogelin long repeat 1\
342-399         #region semenogelin long repeat 2\
414-455         #region semenogelin short repeat 3\
141            #binding_site carbohydate (Asn) (covalent) #status
              absent\
239            #disulfide_bonds interchain #status experimental
SUMMARY        #length 462 #molecular-weight 52117 #checksum 9901

Query Match    84.4%; Score 38; DB 1; Length 462;
Best Local Similarity 85.7%; Pred. No. 1.44e-01;
Matches        6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db    229 EHSSKQV 235
QY    1 EHSSKLQ 7

Search completed: Thu Oct 28 11:25:08 1999
Job time : 16 secs.
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:25:25 1999; MasPar time 2.07 Seconds
Tabular output not generated. 95.556 Million cell updates/sec

Title: >US-09-081-707-10
Description: (1-7) from US09081707.pap
Perfect Score: 45
Sequence: 1 EHSSKIQ 7

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 19.321; Variance 15.201; scale 1.271

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	45	100.0	SEM2_HUMAN	SEMENOGELIN II PRECURS	3.12e-02
2	45	100.0	SEM2_MACMU	SEMENOGELIN II PRECURS	3.12e-02
3	41	91.1	DRP2_HUMAN	DYSTROPHIN-RELATED PRO	5.99e-01
4	40	88.9	Y187_MYCE	HYPOTHETICAL ABC TRANS	1.21e+00
5	39	86.7	KRUH_DROME	KRUEPPEL HOMOLOGOUS PR	2.43e+00
6	39	86.7	GLP1_GLYDI	GLOBIN, MAJOR POLYMERY	2.43e+00
7	39	86.7	PEX2_YARLI	PEROXISOME ASSEMBLY PR	2.43e+00
8	39	86.7	MYSL_DICDI	MYOSIN IJ HEAVY CHAIN.	2.43e+00
9	38	84.4	SEM1_HUMAN	SEMENOGELIN I PROTEIN	4.80e+00
10	38	84.4	YAZ1_SCHPO	HYPOTHETICAL 88.2 KD P	4.80e+00
11	37	82.2	YFD1_ECOLI	HYPOTHETICAL 10.2 KD L	9.35e+00
12	37	82.2	REF1_BACSU	PEPTIDE CHAIN RELEASE	9.35e+00
13	37	82.2	POLN_RVNV	NONSTRUCTURAL POLYPROT	9.35e+00
14	36	80.0	GLP3_GLYDI	GLOBIN, POLYMERIC COMP	1.79e+01
15	36	80.0	RS2_DICDI	40S RIBOSOMAL PROTEIN	1.79e+01
16	36	80.0	OCCR_AGR16	REGULATORY PROTEIN OCC	1.79e+01
17	36	80.0	ATPG_HELVE	ATP SYNTHASE GAMMA CHA	1.79e+01
18	36	80.0	Y070_MYCTU	HYPOTHETICAL 37.3 KD P	1.79e+01
19	36	80.0	FWO3_RABIT	DIMETHYLANILINE MONOX	1.79e+01
20	36	80.0	VTER_HSV11	PROBABLE DNA PACKAGING	1.79e+01
21	36	80.0	Z1P1_YEAST	SYNAPTONEMAL COMPLEX P	1.79e+01
22	36	80.0	YBE7_YEAST	HYPOTHETICAL 150.8 KD	1.79e+01
23	36	80.0	DYHC_ANTCR	DYNEIN BETA CHAIN, CIL	1.79e+01

24	35	77.8	212	1	PRL1_OREMO	PROLACTIN I PRECURSOR	3.37e+01
25	35	77.8	227	1	IPPL1_SCHPO	ISOPENTENYL-DIPHOSPHAT	3.37e+01
26	35	77.8	533	1	CAG2_HUMAN	BETA-1.4 N-ACETYLGLAC	3.37e+01
27	35	77.8	752	1	HPRI1_YEAST	HPRI PROTEIN.	3.37e+01
28	35	77.8	833	1	VIRA_AGR15	WIDE HOST RANGE VIRAL P	3.37e+01
29	35	77.8	952	1	IF41_YEAST	EUKARYOTIC INITIATION	3.37e+01
30	35	77.8	1005	1	P532_HUMAN	P53-BINDING PROTEIN 53	3.37e+01
31	35	77.8	1111	1	KIPL1_YEAST	KINESIN-LIKE PROTEIN K	3.37e+01
32	35	77.8	1139	1	Y060_CAEEL	HYPOTHETICAL 128.6 KD	3.37e+01
33	35	77.8	1597	1	CITRO_MOUSE	CITRON PROTEIN.	3.37e+01
34	35	77.8	1716	1	REPL1_RAT	DNA-DIRECTED RNA POLYM	3.37e+01
35	35	77.8	1717	1	REPL1_MOUSE	DNA-DIRECTED RNA POLYM	3.37e+01
36	35	77.8	2004	1	MOZ_HUMAN	MONOCYTIC LEUKEMIA ZIN	3.37e+01
37	35	77.8	2067	1	BIMB_EMEVI	CELL DIVISION-ASSOCIAT	3.37e+01
38	35	77.8	2431	1	POLN_SFV	NONSTRUCTURAL POLYPROT	3.37e+01
39	35	77.8	2514	1	POLN_ONNVG	NONSTRUCTURAL POLYPROT	3.37e+01
40	35	77.8	3084	1	LMAL_MOUSE	LAMININ ALPHA-1 CHAIN	3.37e+01
41	34	75.6	69	1	Y7K7_BPP22	HYPOTHETICAL 7.7 KD PR	6.25e+01
42	34	75.6	442	1	YAJB_SCHPO	HYPOTHETICAL 51.1 KD PR	6.25e+01
43	34	75.6	1049	1	VP39_YEAST	VACUOLAR ASSEMBLY PROT	6.25e+01
44	34	75.6	1524	1	Y133_HUMAN	HYPOTHETICAL PROTEIN K	6.25e+01
45	34	75.6	2014	1	YU07_YEAST	HYPOTHETICAL 229.9 KD	6.25e+01

ALIGNMENTS

RESULT 1 SEM2_HUMAN STANDARD; PRT; 582 AA.
AC Q02383;

DT 01-JUL-1993 (REL. 26, CREATED)

DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE SEMENOGELIN II PRECURSOR (SGII).

GN SEMG2.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-SEMINAL VESICLE;

RX MEDLINE; 92262479.

RA LUNDWALL A., LILJA H.;

RT "Molecular cloning of epididymal and seminal vesicular transcripts

encoding a semenogelin-related protein."

RL PROC. NATL. ACAD. SCI. U.S.A. 89:4559-4563(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92388176.

RA ULVSBACEK M., LAZURE C., LILJA H., SPURR N.K., RAO V.V., LOEFFLER C.,

HANSMANN I., LUNDWALL A.;

RT "Gene structure of semenogelin I and II. The predominant proteins in

human semen are encoded by two homologous genes on chromosome 20.;"

RL J. BIOL. CHEM. 267:18080-18084(1992).

CC -I- FUNCTION: THREE SEMENOGELIN PROTEINS ARE FOUND IN HUMAN SEMEN,

OF WHICH THE 52 KD SEMENOGELIN I IS THE MOST ABUNDANT. THEY

PARTICIPATE IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE

ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS

OF SEMENOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY

CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS

THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE

PROTEASE.

CC -I- SUBUNIT: DISULFIDE-LINKED COMPLEXES OF SEMENOGELIN I AND THE OTHER

SEMENOGELIN PROTEINS, THE 71- AND 76-KD POLYPEPTIDES.

CC -I- TISSUE SPECIFICITY: SEMINAL VESICLES, AND TO A MUCH LESSER

EXTENT, EPIDIDYMIS.

CC -I- PTM: SEMENOGELIN II IS THOUGHT TO FORM BOTH THE 71-KD POLYPEPTIDE

AND, IN ITS GLYCOSYLATED FORM, THE 76-KD POLYPEPTIDE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZMANN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium";
RL SCIENCE 270:397-403(1995).
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U39696; G1045872; -
DR TIGR; MG187; -
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PFAM; PF00005; ABC_tran; 1.
KW HYPOTHETICAL PROTEIN; ATP-BINDING; POTENTIAL.
FT NP_BIND 40 47
SQ SEQUENCE 585 AA: 65949 MW: EC78C4B8 CRC32;

Query Match 88.9%; Score 40; DB 1; Length 585;
Best Local Similarity 71.4%; Pred. No. 1.21e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 225 EHNAKLO 231
QY 1 EHSSKLQ 7
II::III
1 EHSSKLQ 7

RESULT 5
ID KRUH_DROME STANDARD; PRT; 79 AA.
AC P08155;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE KRUEPPEL HOMOLOGOUS PROTEIN (FRAGMENT).
GN KR-H.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87051757.
RA SCHUH R., AICHER W., GAUL U., COTE S., PREISS A., MAIER D.,
RA SEIFERT E., NAUBER U., SCHRODER C., KEMLER R., JACKLE H.;
RT "A conserved family of nuclear proteins containing structural
RT elements of the finger protein encoded by Kruppel, a Drosophila
RT segmentation gene";
RL CELL 47:1025-1032(1986).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14940; G157796; -
DR FLYBASE; FBgn0001326; Kr-h.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 2.

DR PFAM; PF00096; zf-C2H2; 2.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT.
FT NON_TER 1
FT DOMAIN <1 >79 ZINC-FINGERS.
FT ZN_FING <1 5 C2H2-TYPE.
FT ZN_FING 11 33 C2H2-TYPE.
FT ZN_FING 39 61 C2H2-TYPE.
FT ZN_FING 67 >79 C2H2-TYPE.
FT NON_TER 79
SQ SEQUENCE 79 AA: 9544 MW: 3B8484AF CRC32;

Query Match 86.7%; Score 39; DB 1; Length 79;
Best Local Similarity 71.4%; Pred. No. 2.43e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 49 EHSKILH 55
QY 1 EHSSKLQ 7
III::III
1 EHSSKLQ 7

RESULT 6
ID GLPI_GLYDI STANDARD; PRT; 147 AA.
AC P23216;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE GLOBIN, MAJOR POLYMERIC COMPONENT PL.
OS GLYCERA DIBRANCHIATA (BLOODWORM).
OC EUKARYOTA; METAZOA; ANNELIDA; POLYCHAETA; PHYLLODODA; GLYCERIDAE;
OC GLYCERA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-27; 98-102 AND 119-145.
RX MEDLINE; 91091412.
RA ZAFAR R.S., CHOW L.H., STERN M.S., VINOGRADOV S.N., WALZ D.A.;
RT "The heterogeneity of the polymeric intracellular hemoglobin of
RT Glycera dibranchiata and the cDNA-derived amino acid sequence of one
RT component";
RL BIOCHIM. BIOPHYS. ACTA 1041:117-122(1990).
CC -!- SUBUNIT: POLYMER.
CC -!- THIS PROTEIN IS ONE OF AT LEAST SIX COMPONENTS IN THE POLYMERIC
CC FRACTION OF GLYCERA HEMOGLOBIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54060; G9347; -
DR PIR; S13157; S13157.
DR PROSITE; PS01033; GLOBIN; 1.
DR PFAM; PF00042; globin; 1.
DR HSPF; P02216; 1HEG.
KW HEME; OXYGEN TRANSPORT; RESPIRATORY PROTEIN.
SQ SEQUENCE 147 AA: 16019 MW: 61C9888B CRC32;

Query Match 86.7%; Score 39; DB 1; Length 147;
Best Local Similarity 100.0%; Pred. No. 2.43e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 57 HSSKLQ 62
QY 2 HSSKLQ 7
IIIIII
2 HSSKLQ 7

RESULT 7
ID PEX2_YARLI STANDARD; PRT; 380 AA.
AC Q99155;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PEROXISOME ASSEMBLY PROTEIN PAYS (PEROXIN-2).
GN PEX2 OR PAYS.
OS YARROWIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;
CC DIPODASCACEAE; YARROWIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96355353.
RA EITZEN G.A., TIFORENKO V.I., SMITH J.J., VEENHUIS M.,
RA STILARD R.K., RACHUBINSKI R.A.;
RT "The Yarrowia lipolytica gene PAYS encodes a peroxisomal integral
RT membrane protein homologous to the mammalian peroxisome assembly
RT factor PAF-1.";
RL J. BIOL. CHEM. 271:20300-20306(1996).
CC -!- FUNCTION: REQUIRED FOR THE IMPORT OF SEVERAL PROTEINS INTO
CC PEROXISOMES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL
CC (POTENTIAL).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U43081; G1209744;
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR PFAM; PF00097; 2f-C3HC4; 1.
KW TRANSMEMBRANE; PEROXISOME; ZINC-FINGER.
FT TRANSMEM 153 173
FT TRANSMEM 153 173
FT ZN_FING 265 322
FT ZN_FING 265 322
FT DOMAIN 277 280
FT DOMAIN 354 380
FT DOMAIN 354 380
FT DOMAIN 360 365
FT DOMAIN 367 372
FT DOMAIN 367 372
FT POLY-GLU.
FT POLY-ASP.
FT POLY-ASP.
SQ SEQUENCE 380 AA; 42727 MW; 489F7120 CRC32;
Query Match 86.7%; Score 39; DB 1; Length 380;
Best Local Similarity 100.0%; Pred. No. 2.43e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 92 HSSKIQ 97
QY 2 HSSKIQ 7
RESULT 8
ID MYSJ_DICDI STANDARD; PRT; 2245 AA.
AC P54697;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DE SEMENOGELIN I PROTEIN PRECURSOR (SGI) [CONTAINS: SEMINAL BASIC
DE PROTEIN: ALPHA-INHIBIN-92; ALPHA-INHIBIN-31].
GN SEMGL OR SEMG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RA HAMMER J.A. III, JUNG G.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RA PETERSON M.D., TITUS M.A.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE; 95023928.

RA TITUS M.A., KUSPA A., LOOMIS W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL PROC. NATL. ACAD. SCI. U.S.A. 91:9446-9450(1994).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO CLASS-5 MYOSINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U42409; G1150766;
DR EMBL; L35322; G1039361;
DR DICTYDB; D001095; MYOJ.
DR PFAM; PF00063; myosin_head; 2.
DR PFAM; PF00612; IQ; 3.
DR HSSP; P08799; 1MND.
KW MYOSIN; REPEAT; ATP-BINDING; CALMODULIN-BINDING; ACTIN-BINDING;
KW HEPTAD REPEAT PATTERN; COILED COIL.
FT DOMAIN 1 ?
FT DOMAIN 1 ?
FT DOMAIN 1 ?
FT NP_BIND 174 181
FT NP_BIND 174 181
FT DOMAIN 669 749
FT CONFLICT 191 191
FT CONFLICT 284 284
FT CONFLICT 291 291
FT CONFLICT 332 347
FT CONFLICT 550 550
FT CONFLICT 865 866
FT SEQUENCE 2245 AA; 258478 MW; 48F462D8 CRC32;
SQ SEQUENCE 2245 AA; 258478 MW; 48F462D8 CRC32;
Query Match 86.7%; Score 39; DB 1; Length 2245;
Best Local Similarity 71.4%; Pred. No. 2.43e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 1246 DHKSKIQ 1252
QY 1 EHSKIQ 7
RESULT 9
ID SEM1_HUMAN STANDARD; PRT; 462 AA.
AC P04279; 1987 (REL. 04, CREATED)
DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SEMENOGELIN I PROTEIN PRECURSOR (SGI) [CONTAINS: SEMINAL BASIC
DE PROTEIN: ALPHA-INHIBIN-92; ALPHA-INHIBIN-31].
GN SEMGL OR SEMG.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89109215.
RA LILJA H., ABRAHAMSSON P.-A., LUNDWALL A.;
RT "Semenogelin, the predominant protein in human semen. Primary
RT structure and identification of closely related proteins in the male
RT accessory sex glands and on the spermatozoa.";
RL J. BIOL. CHEM. 264:1894-1900(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE; 92388176.
RA ULVSBACKE M., LAZURE C., LILJA H., SPURR N.K., RAO V.V., LOEFFLER C.,

RA HANSMANN I., LUNDWALL A.;
 RT "Gene structure of semenogelin I and II. The predominant proteins in
 RT human semen are encoded by two homologous genes on chromosome 20";
 RL J. BIOL. CHEM. 267:18080-18084(1992).
 RN [3]
 RP SEQUENCE OF 108-159.
 RX MEDLINE: 85127550.
 RA LILJA H., JEFFERSON J.-O.;
 RT "Amino acid sequence of the predominant basic protein in human
 RT seminal plasma";
 RL FEBS LETT. 182:181-184(1985).
 RN [4]
 RP SEQUENCE OF 108-138.
 RX MEDLINE: 84132557.
 RA SEIDAH N.G., RAMASHARMA K., SAIRAM M.R., CHRETIEN M.;
 RT "Partial amino acid sequence of a human seminal plasma peptide with
 RT inhibin-like activity";
 RL FEBS LETT. 167:98-102(1984).
 RN [5]
 RP SEQUENCE OF 68-159.
 RX MEDLINE: 85216629.
 RA LI C.H., HAMMONDS R.G., RAMASHARMA K., CHUNG D.;
 RT "Human seminal alpha inhibins: isolation, characterization, and
 RT structure";
 RL PROC. NATL. ACAD. SCI. U.S.A. 82:4041-4044(1985).
 CC -!- FUNCTION: SEMENOGELIN I IS THE PREDOMINANT PROTEIN IN HUMAN SEMEN.
 CC IT PARTICIPATES IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE
 CC ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS
 CC OF SEMENOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY
 CC CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS
 CC THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE
 CC PROTEASE.
 CC -!- FUNCTION: ALPHA-INHIBIN-92 AND ALPHA-INHIBIN-31, DERIVED FROM
 CC THE PROTEOLYTIC DEGRADATION OF SEMENOGELIN, INHIBIT THE SECRETION
 CC OF PITUITARY FOLLICLE-STIMULATING HORMONE.
 CC -!- SUBUNIT: OCCURS IN DISULFIDE-LINKED COMPLEXES WHICH MAY ALSO
 CC CONTAIN TWO LESS ABUNDANT 71- AND 76-KD SEMENOGELIN-RELATED
 CC POLYPEPTIDES.
 CC -!- TISSUE SPECIFICITY: SEMINAL VESICLE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04440; G338019; -;
 DR EMBL: 247556; E133812; -;
 DR EMBL: M81650; G487420; -;
 DR PIR: A03254; WTHUB.
 DR PIR: A31489; A31489.
 DR MIM: 182140; -;
 KW SEMEN; SEMINAL VESICLE; REPEAT; SIGNAL; POLYMORPHISM.
 FT SIGNAL 1 23
 FT CHAIN 24 462
 FT MOD_RES 24 24
 FT PEPTIDE 68 159
 FT PEPTIDE 108 138
 FT PEPTIDE 108 159
 FT DISULFID 239 239
 FT REPEAT 174 215
 FT REPEAT 234 275
 FT REPEAT 282 339
 FT REPEAT 342 399
 FT REPEAT 414 455
 FT REPEAT 79 79
 FT VARIANT 423 423
 FT CONFLICT 423 423
 FT SEQUENCE 462 AA; 52131 MW; C87515C7 CRC32;
 Query Match 84.4%; Score 38; DB 1; Length 462;
 Best Local Similarity 85.7%; Pred. No. 4.80e+00;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 229 EHSSKQV 235
 QY 1 EHSSKLQ 7
 RESULT 10
 ID YA21 SCHPO STANDARD; PRT; 777 AA.
 AC 009697;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DE HYPOTHEICAL 88.2 KD PROTEIN C2F7.18C IN CHROMOSOME I.
 GN SPAC2F7 18C OR SPAC13A11.01C
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
 OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE OF 1-147 FROM N.A.
 RC STRAIN-972;
 RA HUNT S., DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A.,
 RA WALSH S.V.;
 RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE OF 114-777 FROM N.A.
 RC STRAIN-972;
 RA GENTLES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
 RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 254096; G984222; -;
 DR EMBL: 250142; G1052801; -;
 DR PFAM: PF00610; DEP; 1.
 DR PFAM: PF00611; FCH; 1.
 DR PFAM: PF00620; RBOGAP; 1.
 DR HSP: P01888; 1BMG.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 777 AA; 88207 MW; BDCAFB7 CRC32;
 Query Match 84.4%; Score 38; DB 1; Length 777;
 Best Local Similarity 71.4%; Pred. No. 4.80e+00;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 114 DHSKLQ 120
 QY 1 EHSSKLQ 7
 RESULT 11
 ID YPDI_ECOLI STANDARD; PRT; 91 AA.
 AC O32528;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DE HYPOTHEICAL 10.2 KD LIPOPROTEIN IN EVGS-DDG REGION PRECURSOR.
 GN YPDI.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,

```
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.,
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
CC -|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000326; G2367134; -.
DR ECOGENE; EG14376; YPDI.
DR PROSITE; PS00013; PROKAR LIPOPROTEIN: 1.
KW HYPOTHETICAL PROTEIN; MEMBRANE; LIPOPROTEIN; SIGNAL.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 91
FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 91 AA; 10162 MW; 294E168F CRC32;
Query Match 82.2%; Score 37; DB 1; Length 91;
Best Local Similarity 83.3%; Pred. No. 9.35e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 48 HSTKLQ 53
QY 2 HSSKLQ 7
RESULT 12
ID RF1_BACSU STANDARD; PRT; 356 AA.
AC P45872;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PEPTIDE CHAIN RELEASE FACTOR 1 (RF-1).
GN PRFA.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA GLASER P., DE LA FUENTE V., DANCHIN A.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- FUNCTION: PEPTIDE CHAIN RELEASE FACTOR 1 DIRECTS THE TERMINATION
CC OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN TERMINATION
CC CODONS UAG AND UAA.
CC -|- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
CC FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49782; G853776; -.
DR EMBL; Z99122; E1184607; -.
DR SUBTILIST; BG11244; PRFA.
DR PROSITE; PS00745; RF_PROK_I; 1.
DR PFAM; PF00472; RF-1; 1.
KW PROTEIN BIOSYNTHESIS.
SQ SEQUENCE 356 AA; 40233 MW; 7BF8DDDD CRC32;
Query Match 82.2%; Score 37; DB 1; Length 356;
```

```
Best Local Similarity 57.1%; Pred. No. 9.35e+00;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 346 DQASKLQ 352
QY 1 EHSSKLQ 7
RESULT 13
ID POLN_RRVN STANDARD; PRT; 2479 AA.
AC P13887;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL POLYPROTEIN [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO
DE NSP4].
OS ROSS RIVER VIRUS (STRAIN NB5092) (RRV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
OC ALPHAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88179556.
RA FARAGHER S.G., MEEK A.D.J., RICE C.M., DALGARNO L.;
RT "Genome sequences of a mouse-avirulent and a mouse-virulent strain of
RT Ross River virus.";
RL VIROLOGY 163:509-526(1988).
CC -|- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
CC -|- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -|- READTHROUGH OF THE TERMINATOR CODON UGA OCCURS BETWEEN THE
CC CODONS FOR 1862-PHE AND 1863-LEU.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20162; G1256538; ALT_FRAME.
DR PIR; A28605; MNWVRA.
KW POLYPROTEIN; NONSTRUCTURAL PROTEIN; RNA-BINDING; HELICASE.
FT CHAIN 1 533 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 534 1331 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1332 1868 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1869 2479 NONSTRUCTURAL PROTEIN NSP4.
SQ SEQUENCE 2479 AA; 276436 MW; CF5E6779 CRC32;
Query Match 82.2%; Score 37; DB 1; Length 2479;
Best Local Similarity 71.4%; Pred. No. 9.35e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 1232 DHSMKLQ 1238
QY 1 EHSSKLQ 7
RESULT 14
ID GLP3_GLYDI STANDARD; PRT; 147 AA.
AC P21660;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)
DE GLOBIN, POLYMERIC COMPONENT P3.
OC GLYCERA DIBRANCHIATA (BLOODWORM).
OS EUKARYOTA; METAZOA; ANNELIDA; POLYCHAETA; PHYLLODOCIDA; GLYCERIDAE;
OC GLYCERA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 91072390.
RA ZAFAR R.S., CHOW L.H., STERN M.S., SCULLY J.S., SHARMA P.R.,
RA VINOGRADOV S.N., WALZ D.A.;
RT "The cDNA sequences encoding two components of the polymeric fraction
```

RT of the intracellular hemoglobin of Glycera dibranchiata.";
RL J. BIOL. CHEM. 265:21843-21851(1990).
CC -!- SUBUNIT: POLYMER.
CC -!- THIS PROTEIN IS ONE OF AT LEAST SIX COMPONENTS IN THE POLYMERIC
CC FRACTION OF GLYCERA HEMOGLOBIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55444; G159147; -
DR PIR; B36529; B36529
DR PROSITE; PS01033; GLOBIN; 1.
DR PFAM; PF00042; globin; 1.
DR HSP; P02216; 1HGB.
KW HEME; OXYGEN TRANSPORT; RESPIRATORY PROTEIN.
SQ SEQUENCE 147 AA; 15977 MW; 747D26B7 CRC32;

Query Match 80.0%; Score 36; DB 1; Length 147;
Best Local Similarity 83.3%; Pred. No. 1.79e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 57 HNSKLQ 62
QY 1:||||
2 HSSKLQ 7

RESULT 15
ID RS2 DICDI STANDARD; PRT; 265 AA.
AC P27685;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE 40S RIBOSOMAL PROTEIN S2 (S4) (LLREP3 PROTEIN).
GN RPS2.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE; 91319544.
RA PROFFITT J.A., JAGGER P.S., WILSON G.A., DONOVAN J.T.J.,
RA WIDDOWSON D.C.C., HAMES B.D.;
RT "A developmentally regulated gene encodes the dictyostelium homolog
RT of yeast ribosomal protein S4 and mammalian LLRep3 proteins.";
RL NUCLEIC ACIDS RES. 19:3867-3873(1991).
CC -!- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56297; G7299; -
DR PIR; S22297; S22297.
DR DICTYDB; DD01039;
DR PROSITE; PS00585; RIBOSOMAL_S5; 1.
DR PFAM; PF00333; S5; 1.
DR HSP; P02357; 1PKP.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 265 AA; 28717 MW; 7A26BB45 CRC32;

Query Match 80.0%; Score 36; DB 1; Length 265;
Best Local Similarity 83.3%; Pred. No. 1.79e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 255 EHSACL 260
|||:|
QY 1 EHSSKL 6

Search completed: Thu Oct 28 11:25:34 1999
Job time : 9 secs.

THIS PAGE BLANK (USPTO)

WPSRELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:25:51 1999; MasPar time 4.17 Seconds
Tabular output not generated. 91.547 Million cell updates/sec

Title: >US-09-081-707-10
Description: (1-7) from US09081707.pep
Perfect Score: 45
Sequence: 1 EHSSKLIQ 7

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spstreml9
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 18.462; Variance 16.244; scale 1.136

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	42	93.3	509	10	F6N15.5 PROTEIN.	8.55e-01
2	42	93.3	627	4	GUANYLATE KINASE ASSOC	8.55e-01
3	42	93.3	666	11	GUANYLATE KINASE ASSOC	8.55e-01
4	42	93.3	675	4	DAP-1 BETA.	8.55e-01
5	42	93.3	692	11	PSD-95 BINDING PROTEIN	8.55e-01
6	42	93.3	977	4	DAP-1 ALPHA.	8.55e-01
7	42	93.3	992	11	PSD-95/SAP90-ASSOCIATE	8.55e-01
8	42	93.3	1548	5	SIMILARITY TO HUMAN PR	8.55e-01
9	42	93.3	1638	5	GUANINE NUCLEOTIDE EXC	8.55e-01
10	42	93.3	2488	5	GUANINE NUCLEOTIDE EXC	8.55e-01
11	40	88.9	265	10	OEC 23KD PROTEIN.	3.23e+00
12	40	88.9	849	2	85KDA STRAIN-SPECIFIC	3.23e+00
13	40	88.9	1795	2	AVIRULENCE PROTEIN.	3.23e+00
14	39	86.7	585	2	FROM BASES 1472112 TO	6.19e+00
15	38	84.4	186	4	DYNACTIN SUBUNIT.	1.17e+01
16	38	84.4	210	2	MAJOR SURFACE PROTEIN	1.17e+01
17	38	84.4	224	2	MG2+ TRANSPORT PROTEIN	1.17e+01
18	38	84.4	273	14	EARLY-EXPRESSED PROTEIN	1.17e+01
19	38	84.4	373	4	POTENTIAL TRANSCRIPTION	1.17e+01
20	38	84.4	402	5	T27E7.3 PROTEIN.	1.17e+01

21	38	84.4	420	2	Q51709	CYTOCHROME BA(3) PRECU	1.17e+01
22	38	84.4	556	14	O36369	TEGUMENT PROTEIN.	1.17e+01
23	38	84.4	617	10	O80686	T3K9.25 PROTEIN.	1.17e+01
24	38	84.4	656	2	O87712	HSP70.	1.17e+01
25	38	84.4	747	11	O55007	TULIP 1.	1.17e+01
26	38	84.4	866	11	O55008	TULIP 2.	1.17e+01
27	38	84.4	1244	5	O21778	MITOTIC CHROMOSOME AND	1.17e+01
28	38	84.4	1464	5	O23995	TOLLOID RELATED-1.	1.17e+01
29	38	84.4	1978	4	O15154	TRIP230.	1.17e+01
30	37	82.2	278	2	Q40797	WAS F268 AND F157P BEF	2.18e+01
31	37	82.2	378	8	O24004	CYTOCHROME B.	2.18e+01
32	37	82.2	378	8	O21145	CYTOCHROME B.	2.18e+01
33	37	82.2	587	5	O02427	FORHEAD HOMOLOG.	2.18e+01
34	37	82.2	604	3	O12166	VOR3227W FROM CHROMOSO	2.18e+01
35	37	82.2	700	13	Q91453	STONUSTOXIN BETA-SUBUN	2.18e+01
36	37	82.2	708	13	Q98993	VERRUCOTOXIN ALPHA PRE	2.18e+01
37	37	82.2	813	11	P70625	ZONULA OCCLUDENS 2 PRO	2.18e+01
38	37	82.2	883	10	O04024	RNA HELICASE.	2.18e+01
39	37	82.2	950	10	O04024	HYPOTHETICAL 108.3 KD	2.18e+01
40	37	82.2	1339	10	O82180	T4C15.22 PROTEIN.	2.18e+01
41	36	80.0	290	2	O85597	2-OXOGLUTARIC DEHYDROG	4.02e+01
42	36	80.0	539	2	O85280	50KDA STRAIN-SPECIFIC	4.02e+01
43	36	80.0	734	14	P89438	HERPES SIMPLEX VIRUS T	4.02e+01
44	36	80.0	920	5	Q94893	89B HELICASE (FRAGMENT	4.02e+01
45	36	80.0	3507	5	Q23587	CODED FOR BY C. ELEGAN	4.02e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	509 AA.
ID O81317			
AC O81317			
DT 01-NOV-1998	(TREMBLER. 08, CREATED)		
DT 01-NOV-1998	(TREMBLER. 08, LAST SEQUENCE UPDATE)		
DT 01-NOV-1998	(TREMBLER. 08, LAST ANNOTATION UPDATE)		
DE F6N15.5 PROTEIN.			
GN F6N15.5			
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).			
OC EUKARYOTA; VIRIDIPHYTES; EMBRYOPHYTA; TRACHEOPHYTA;			
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;			
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=CV. COLUMBIA;			
RA WASHU;			
RT "The A. thaliana Genome Sequencing Project.";			
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=CV. COLUMBIA;			
RA RYAN E., EDWARDS J., PAPE K.;			
RT "The sequence of A. thaliana F6N15.";			
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
RN [3]			
RP SEQUENCE FROM N.A.			
RC STRAIN=CV. COLUMBIA;			
RA WATERSTON R.;			
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR EMBL; AF069299; G3193321;			
SQ SEQUENCE 509 AA; 58562 MW; 5A682D0D CRC32;			
Query Match	93.3%;	Score 42;	DB 10; Length 509;
Best Local Similarity	85.7%;	Pred. No. 8.55e-01;	
Matches	6;	Conservative	1; Mismatches 0; Indels 0; Gaps 0;
Db	375 EHSPKIQ 381		
Qy	1 EHSSKLIQ 7		
RESULT 2			
ID P78335	PRELIMINARY;	PRT;	627 AA.
AC P78335;			

Mon Nov 1 12:18:26 1999

DT 01-MAY-1997 (TREMREL. 03, CREATED)
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE GUANYLATE KINASE ASSOCIATED PROTEIN.
 GN GKAP.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 97177144.
 RX KIM E., NAISBITT S., HSUEH Y.P., RAO A., ROTHSCHILD A., CRAIG A.M.,
 RA SHENG M.,
 RA "GKAP, a novel synaptic protein that interacts with the guanylate
 RT kinase-like domain of the PSD-95/SAP90 family of channel clustering
 RT molecules.";
 RL J. CELL BIOL. 136:669-678(1997).
 DR EMBL; U67988; G1857137; -.
 SQ SEQUENCE 627 AA; 70051 MW; C40A3CBB CRC32;

 Query Match 93.3%; Score 42; DB 4; Length 627;
 Best Local Similarity 85.7%; Pred. No. 8.55e-01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 Db 93 EHSKQLQ 99
 QY 1 EHSSKQLQ 7

 RESULT 3
 ID P97841 PRELIMINARY; PRT; 666 AA.
 AC P97841;
 DT 01-MAY-1997 (TREMREL. 03, CREATED)
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE GUANYLATE KINASE ASSOCIATED PROTEIN.
 GN GKAP.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 97177144.
 RX KIM E., NAISBITT S., HSUEH Y.P., RAO A., ROTHSCHILD A., CRAIG A.M.,
 RA SHENG M.,
 RA "GKAP, a novel synaptic protein that interacts with the guanylate
 RT kinase-like domain of the PSD-95/SAP90 family of channel clustering
 RT molecules.";
 RL J. CELL BIOL. 136:669-678(1997).
 DR EMBL; U67987; G1857139; -.
 SQ SEQUENCE 666 AA; 74233 MW; B3E3BB93 CRC32;

 Query Match 93.3%; Score 42; DB 11; Length 666;
 Best Local Similarity 85.7%; Pred. No. 8.55e-01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 Db 103 EHSKQLQ 109
 QY 1 EHSSKQLQ 7

 RESULT 4
 ID O14489 PRELIMINARY; PRT; 675 AA.
 AC O14489;
 DT 01-JAN-1998 (TREMREL. 05, CREATED)
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE DAP-1 BETA.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 97177144.
 RX SATOH K., YANAI H., SENDA T., KOHU K., NAKAMURA T., OKUMURA N.,
 RA MATSUMINE A., KOBAYASHI S., TOYOSHIMA K., AKIYAMA T.,
 RA "DAP-1, a novel protein that interacts with the guanylate kinase-like
 RT domains of hDLG and PSD-95.";
 RL GENES TO CELLS 2:415-424(1997).
 DR EMBL; AB000277; D1024135; -.
 SQ SEQUENCE 977 AA; 108872 MW; FFF0A0A0 CRC32;

 Query Match 93.3%; Score 42; DB 4; Length 977;
 Best Local Similarity 85.7%; Pred. No. 8.55e-01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 395 EHSKQLQ 401

RX MEDLINE; 97431353.
 RA SATOH K., YANAI H., SENDA T., KOHU K., NAKAMURA T., OKUMURA N.,
 RA MATSUMINE A., KOBAYASHI S., TOYOSHIMA K., AKIYAMA T.,
 RA "DAP-1, a novel protein that interacts with the guanylate kinase-like
 RT domains of hDLG and PSD-95.";
 RL GENES TO CELLS 2:415-424(1997).
 DR EMBL; AB000276; D1024134; -.
 SQ SEQUENCE 675 AA; 75260 MW; 1FE2AC4 CRC32;

 Query Match 93.3%; Score 42; DB 4; Length 675;
 Best Local Similarity 85.7%; Pred. No. 8.55e-01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 Db 93 EHSKQLQ 99
 QY 1 EHSSKQLQ 7

 RESULT 5
 ID O54773 PRELIMINARY; PRT; 592 AA.
 AC O54773;
 DT 01-JUN-1998 (TREMREL. 06, CREATED)
 DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE PSD-95 BINDING PROTEIN.
 GN RATTUS NORVEGICUS (RAT).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 98089008.
 RX KAWASHIMA N., TAKAMIYA K., SUN J., KITABATAKE A., SOBUE K.,
 RT "Differential expression of isoforms of PSD-95 binding protein
 RT (GKAP/SAPAP1) during rat brain development.";
 RL FEBS LETT. 418:301-304(1997).
 DR EMBL; AB003594; D1025176; -.
 SQ SEQUENCE 592 AA; 76991 MW; CD061633 CRC32;

 Query Match 93.3%; Score 42; DB 11; Length 692;
 Best Local Similarity 85.7%; Pred. No. 8.55e-01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 Db 101 EHSKQLQ 107
 QY 1 EHSSKQLQ 7

 RESULT 6
 ID O14490 PRELIMINARY; PRT; 977 AA.
 AC O14490;
 DT 01-JAN-1998 (TREMREL. 05, CREATED)
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE DAP-1 ALPHA.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 97431353.
 RX SATOH K., YANAI H., SENDA T., KOHU K., NAKAMURA T., OKUMURA N.,
 RA MATSUMINE A., KOBAYASHI S., TOYOSHIMA K., AKIYAMA T.,
 RA "DAP-1, a novel protein that interacts with the guanylate kinase-like
 RT domains of hDLG and PSD-95.";
 RL GENES TO CELLS 2:415-424(1997).
 DR EMBL; AB000277; D1024135; -.
 SQ SEQUENCE 977 AA; 108872 MW; FFF0A0A0 CRC32;

 Query Match 93.3%; Score 42; DB 4; Length 977;
 Best Local Similarity 85.7%; Pred. No. 8.55e-01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 395 EHSKQLQ 401


```
|||||
Qy 1 EHSSKLQ 7

Best Local Similarity 85.7%; Pred. No. 8.55e-01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
ID P97836 PRELIMINARY; PRT; 992 AA.
AC P97836;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PSD-95/SAP90-ASSOCIATED PROTEIN-1.
OS RATUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SC SCTROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RA TAKAI Y., TAKEUCHI M., IRIE M., HATA Y.;
RL J. BIOL. CHEM. 0:0-0(0).
DR EMBL; U67137; G1864087; -.
SQ SEQUENCE 992 AA; 110177 MW; 7D090FDD CRC32;

Query Match 93.3%; Score 42; DB 11; Length 992;
Best Local Similarity 85.7%; Pred. No. 8.55e-01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 401 EHSPKLQ 407
|||||
Qy 1 EHSSKLQ 7

RESULT 8
ID P91339 PRELIMINARY; PRT; 1548 AA.
AC P91339;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILARITY TO HUMAN PROTO-ONCOGENE DBL.
CN F55C7.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA CRAXTON M., BURTON J., CONNELL M., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SNALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RN NATURE 368:32-38(1994).
RP SEQUENCE FROM N.A.
RA DU Z., LE T.;
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RC STRAIN-BRISTOL N2;
RA SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U80436; G1703564; -.
DR PFAM; PF00621; RHOGEF; 1.
SQ SEQUENCE 1548 AA; 179071 MW; 63FFFA8A CRC32;

Query Match 93.3%; Score 42; DB 5; Length 1548;

Best Local Similarity 85.7%; Pred. No. 8.55e-01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 468 EHSSRLQ 442
|||||
Qy 1 EHSSKLQ 7

RESULT 9
ID O61529 PRELIMINARY; PRT; 1638 AA.
AC O61529;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR UNC-73B.
CN UNC-73.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA STEVEN R., KUBISESKI T. J., ZHENG H., KULKARNI S., MANCILLAS J.,
RA RUIZ MORALES A., HOGUE C. W. V., PAWSON T., CULOTTI J.;
RT "UNC-73 activates the Rac GTPase and is required for cell and growth
RT cone migrations in C. elegans.";
RL CELL 92:785-795(1998).
DR EMBL; AF048835; G2944398; -.
SQ SEQUENCE 1638 AA; 189071 MW; 83310A21 CRC32;

Query Match 93.3%; Score 42; DB 5; Length 1638;
Best Local Similarity 85.7%; Pred. No. 8.55e-01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 468 EHSSRLQ 474
|||||
Qy 1 EHSSKLQ 7

RESULT 10
ID O61528 PRELIMINARY; PRT; 2488 AA.
AC O61528;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR UNC-73A.
CN UNC-73.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA STEVEN R., KUBISESKI T. J., ZHENG H., KULKARNI S., MANCILLAS J.,
RA RUIZ MORALES A., HOGUE C. W. V., PAWSON T., CULOTTI J.;
RT "UNC-73 activates the Rac GTPase and is required for cell and growth
RT cone migrations in C. elegans.";
RL CELL 92:785-795(1998).
DR EMBL; AF048834; G2944396; -.
SQ SEQUENCE 2488 AA; 282876 MW; 40A8CB92 CRC32;

Query Match 93.3%; Score 42; DB 5; Length 2488;
Best Local Similarity 85.7%; Pred. No. 8.55e-01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 468 EHSSRLQ 474
|||||
Qy 1 EHSSKLQ 7

RESULT 11
```

ID Q40407 PRELIMINARY; PRT: 265 AA.
AC Q40407;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE OEC 23KD PROTEIN.
GN OEC 23.
OS NARCISUS PSEUDONARCISUS (DAFFODIL).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHITES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; ASPARAGALES;
OC AMARYLLIDACEAE; NARCISUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PARACOROLLA;
RX MEDLINE: 96085151.
RA NITZSCHE W., BEYER P.;
RA NITZSCHE W., VANDEKERCKHOVE J., TADROS M.H., LINTIG J.V.,
RT "Carotene desaturation is linked to a respiratory redox pathway in
RT Narcissus pseudonarcissus chromoplast membranes. Involvement of a
RT 23-kDa oxygen-evolving-complex-like protein.";
RL EUR J. BIOCHEM. 233:864-872(1995).
DR EMBL: X78816; G780273; -;
DR MENDEL; 11966; NARPS:psbp.1.
SQ SEQUENCE 265 AA; 28521 MW; BF1116D6 CRC32;

Query Match 88.9%; Score 40; DB 10; Length 265;
Best Local Similarity 71.4%; Pred. No. 3.23e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 35 EHSARLO 41
|:|:|:|
QY 1 EHSSKLQ 7

RESULT 12
ID O85281 PRELIMINARY; PRT: 849 AA.
AC O85281;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 85KDA SPRAIN-SPECIFIC ANTIGEN.
DE EHRlichia RISTICII.
OS BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAE; EHRlichiae; EHRlichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-90-12;
RX MEDLINE: 98339868.
RA BISWAS B., VEMULAPALLI R., DUTTA S.K.;
RT "Molecular basis for antigenic variation of a protective
RT strain-specific antigen of Ehrlichia risticii.";
RL INFECT. IMMUN. 66:3682-3688(1998).
DR EMBL: AF059673; G3414803; -;
SQ SEQUENCE 849 AA; 93299 MW; BC65E036 CRC32;

Query Match 88.9%; Score 40; DB 2; Length 849;
Best Local Similarity 57.1%; Pred. No. 3.23e+00;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 815 DHATKLO 821
|:|:|:|
QY 1 EHSSKLQ 7

RESULT 13
ID O66101 PRELIMINARY; PRT: 1795 AA.
AC O66101;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE AVIRULENCE PROTEIN.
GN AVRE.
OS PSEUDOMONAS SYRINGAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DC3000;
RX MEDLINE: 98115919.
RA BOGDANOV A.J., KIM J.F., WEI Z., KOLCHINSKY P., CHARKOWSKI A.O.,
RA CONLIN A.K., COLLIER A., BEER S.V.;
RT "Homology and functional similarity of an hrp-linked pathogenicity
RT locus, dsrEF, of *Erwinia amylovora* and the avirulence locus avrE of
RT *Pseudomonas syringae* pathovar tomato.";
RL PROC. NATL. ACAD. SCI. U.S.A. 95:1325-1330(1998).
DR EMBL: U97505; G2978503; -;
SQ SEQUENCE 1795 AA; 195351 MW; 1A8A737 CRC32;

Query Match 88.9%; Score 40; DB 2; Length 1795;
Best Local Similarity 85.7%; Pred. No. 3.23e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 575 BQSSKLQ 581
|:|:|:|
QY 1 EHSSKLQ 7

RESULT 14
ID P76092 PRELIMINARY; PRT: 585 AA.
AC P76092;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FROM BASES 1472112 TO 1485151
DE (SECTION 128 OF 400) OF THE COMPLETE GENOME (SECTION 128 OF 400).
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT III G., BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER J.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RT SCIENCE 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RA KITAKAWA M., KASAI H., BABA T., HONJO A., ISONO K.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RX MEDLINE: 92190338.
RA MOSZER I., GLASER P., DANCHIN A.;
RT "Multiple IS insertion sequences near the replication terminus in
RT *Escherichia coli* K-12.";
RL BIOCHIMIE 73:1361-1374(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RX MEDLINE: 95206938.
RA MORIYA H., KASAI H., ISONO K.;
RT "Cloning and characterization of the hrpA gene in the terC region of
RT *Escherichia coli* that is highly similar to the DEAH family RNA
RT helicase genes of *Saccharomyces cerevisiae*.";
RL NUCLEIC ACIDS RES. 23:595-598(1995).
DR EMBL: AE000238; G1787678; -;
DR EMBL: D85081; D1026334; -;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 585 AA; 65455 MW; BC4AD908 CRC32;

Query Match 86.7%; Score 39; DB 2; Length 585;

Best Local Similarity 71.4%; Pred. No. 6.19e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 44 EHSRLQ 50
| | | | |
QY 1 EHSSKLQ 7

RESULT 15
ID 075935 PRELIMINARY; PRT; 186 AA.
AC 075935;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE DYNACTIN SUBUNIT.
GN P22.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389897.
RA KARKI S., LAMONTE B., HOLZBAUR E.L.F.;
RT "Characterization of the p22 subunit of dynactin reveals the
RT localization of cytoplasmic dynein and dynactin to the midbody of
RT dividing cells.";
RL J. CELL BIOL. 142:1023-1034(1998).
DR EMBL; AF082513; G3649857; -.
SQ SEQUENCE 186 AA; 21119 MW; 7B47CEC9 CRC32;

Query Match 84.4%; Score 38; DB 4; Length 186;
Best Local Similarity 57.1%; Pred. No. 1.17e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 121 EHARLQ 127
| | | | |
QY 1 EHSSKLQ 7

Search completed: Thu Oct 28 11:26:30 1999
Job time : 39 secs.

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:29:02 1999; MasPar time 3.25 Seconds
 Tabular output not generated. 45.813 Million cell updates/sec

Title: >US-09-081-707-11
 Description: (1-7) from US09081707.pap
 Perfect Score: 49
 Sequence: 1 QNKISYQ 7

Scoring table: PAM 150
 Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39

Statistics: Mean 14.672; Variance 35.519; scale 0.413

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	49	100.0	287	32	W51096	1.11e+01
2	43	87.8	17	26	W33439	Oligopeptide 19 cleav
3	43	87.8	17	17	R89602	Prostate specific ant
4	42	85.7	7	17	R89612	Prostate specific ant
5	42	85.7	7	26	W33449	Oligopeptide 29 cleav
6	42	85.7	8	26	W33546	Oligopeptide 132 base
7	42	85.7	8	26	W33435	Oligopeptide 15 cleav
8	42	85.7	8	17	R89598	Prostate specific ant
9	42	85.7	8	17	R89712	Prostate specific ant
10	42	85.7	9	17	R89599	Prostate specific ant
11	42	85.7	9	17	R89718	Prostate specific ant
12	42	85.7	9	26	W33552	Oligopeptide 132 base
13	42	85.7	9	26	W33488	Oligopeptide 68 cleav
14	42	85.7	10	26	W33448	Oligopeptide 28 cleav
15	42	85.7	10	26	W33429	Oligopeptide 9 cleav
16	42	85.7	10	26	W33463	Oligopeptide 43 cleav

17	42	85.7	10	26	W33437	Oligopeptide 17 cleav	9.60e+01
18	42	85.7	10	26	W33500	Oligopeptide 80 cleav	9.60e+01
19	42	85.7	10	17	R89626	Prostate specific ant	9.60e+01
20	42	85.7	10	17	R89600	Prostate specific ant	9.60e+01
21	42	85.7	10	17	R89663	Prostate specific ant	9.60e+01
22	42	85.7	10	17	R89592	Prostate specific ant	9.60e+01
23	42	85.7	11	17	R89657	Prostate specific ant	9.60e+01
24	42	85.7	11	26	W33533	Oligopeptide 113 clea	9.60e+01
25	42	85.7	11	17	R89585	Prostate specific ant	9.60e+01
26	42	85.7	11	17	R89699	Prostate specific ant	9.60e+01
27	42	85.7	11	26	W33422	Oligopeptide 2 cleav	9.60e+01
28	42	85.7	12	26	W33493	Oligopeptide 73 cleav	9.60e+01
29	42	85.7	12	26	W33497	Oligopeptide 77 cleav	9.60e+01
30	42	85.7	12	26	W33499	Oligopeptide 80 based	9.60e+01
31	42	85.7	12	26	W33496	Oligopeptide 76 based	9.60e+01
32	42	85.7	12	17	R89662	Prostate specific ant	9.60e+01
33	42	85.7	12	26	W33492	Oligopeptide 72 cleav	9.60e+01
34	42	85.7	12	17	R89658	Prostate specific ant	9.60e+01
35	42	85.7	12	17	R89656	Prostate specific ant	9.60e+01
36	42	85.7	12	17	R89655	Prostate specific ant	9.60e+01
37	42	85.7	12	17	R89652	Prostate specific ant	9.60e+01
38	42	85.7	12	17	R89652	Prostate specific ant	9.60e+01
39	42	85.7	12	26	W33489	Oligopeptide 69 cleav	9.60e+01
40	42	85.7	12	17	R89593	Prostate specific ant	9.60e+01
41	42	85.7	14	26	W33495	Oligopeptide 75 based	9.60e+01
42	42	85.7	25	26	W33425	Oligopeptide 5 cleav	9.60e+01
43	42	85.7	25	17	R89649	Prostate specific ant	9.60e+01
44	42	85.7	462	17	R89583	Human semenogelin I.	9.60e+01
45	42	85.7	462	26	W33420	Human semenogelin I.	9.60e+01

ALIGNMENTS

RESULT 1
 ID W51096 standard; Protein; 287 AA.
 AC W51096;
 DT 14-SEP-1998 (first entry)
 DE Ehrlichia canis VSAL protein (partial sequence).
 KW MAP1 homologue; variable surface antigen; VSAL; rickettsia;
 KN DNA vaccine.
 OS Ehrlichia canis.
 FH Key
 FT Peptide
 FT 1..25
 FT /note= "putative signal peptide"
 FN W09816554-AL
 PD 23-APR-1998;
 PF 17-OCT-1997;
 PR 17-OCT-1996; US-733230.
 PA (UYFL) UNIV FLORIDA.
 PI Barbet AF, Burridge MJ, Ganta RR, McGuire TC,
 PI Nyika A, Rurangirwa FR;
 DR WPI; 98-251232/22.
 DR N-PSDB; V07180.
 PT Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 PS Claim 3; Fig 2C; 39pp; English.
 CC This is the full-length variable surface antigen VSAL protein of
 CC Ehrlichia canis. Its amino acid sequence was deduced from open
 CC reading frame (ORF) 1 of a genomic locus (see V07180) of E. canis
 CC that was obtained on the basis of homology to the major antigenic
 CC protein MAP1 (see W51088) of Cowdria ruminantium. This genomic
 CC locus included 2 ORFs encoding similar, but non-identical proteins
 CC (see W51096-97). A claimed composition comprises a nucleic acid
 CC (see V07176-82) encoding a polypeptide (see W51088-99) that elicits
 CC a protective immune response against a rickettsial pathogen. The
 CC nucleic acid is used, in human or veterinary medicine, in vaccines
 CC to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
 CC species. The Ehrlichia antigenic polypeptides can be used
 CC diagnostically to detect antibodies associated with Ehrlichia
 CC infection (claimed).
 CC Sequence 287 AA;
 SQ

```

Query Match      100.0%; Score 49; DB 32; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.11e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 214 qnkisyq 220
   |||||
QY 1 QNKISYQ 7

RESULT 2
ID W33439 standard; Peptide; 17 AA.
AC W33439;
DT 27-MAR-1998 (first entry)
DE Oligopeptide 19 cleaved by free prostate specific antigen.
KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
KW cleavage site; semenogelin I; chymotrypsin-like; serum PSA;
KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
KW detection; cytotoxic conjugate; activation; quantitative assay.
OS Synthetic.
FH Key Location/Qualifiers
FT Cleavage_site 9..10 /note= "PSA specific cleavage site"
PN W09712624-A1.
PD 10-APR-1997.
PF 02-OCT-1996; U15713.
PR 06-OCT-1995; US-540412.
PA (MERI ) MERCK & CO INC.
PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
PT Oligopeptide(s) recognised and cleaved by free prostate specific
PT antigen - useful in assays for active antigen, and in
PT oligopeptide-drug conjugates for prostatic cancer treatment
PS Disclosure; Page 81; 192pp; English.
CC The present sequence is a novel oligopeptide that is recognised and
CC proteolytically cleaved by free prostate specific antigen (PSA). It was
CC designed based on the sequences surrounding the PSA cleavage sites of
CC semenogelin I, a major sperm entrapping gel protein. PSA, which has
CC chymotrypsin-like specificity is responsible for dissolution of the gel
CC structure formed at ejaculation by proteolysis of the major gel proteins
CC (semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
CC releases progressively motile spermatozoa. PSA complexed to
CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
CC of the prostate. Prostate metastases are also known to secrete
CC immunologically reactive PSA since serum PSA is detectable at high levels
CC in prostatectomised patients showing widespread metastatic prostate
CC cancer. Cytotoxic compounds that could be activated by the proteolytic
CC activity of PSA should also be prostate cell specific as well as specific
CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
CC and this oligopeptide are useful in treatment of prostate cancer. The
CC oligopeptide can also be used in a quantitative assay for enzymatically
CC active PSA.
SQ Sequence 17 AA;

Query Match      87.8%; Score 43; DB 26; Length 17;
Best Local Similarity 85.7%; Pred. No. 7.10e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 3 dnkisyq 9
   :|||||
QY 1 QNKISYQ 7

RESULT 3
ID R89602 standard; Peptide; 17 AA.
AC R89602;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
KW gel structure; dissolution; prostate specific antigen; proteolysis;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
SQ Sequence 17 AA;

Query Match      87.8%; Score 43; DB 17; Length 17;
Best Local Similarity 85.7%; Pred. No. 7.10e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 3 dnkisyq 9
   :|||||
QY 1 QNKISYQ 7

RESULT 4
ID R89612 standard; peptide; 7 AA.
AC R89612;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
KW gel structure; dissolution; prostate specific antigen; proteolysis;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 6..7 /note= "prostate specific antigen proteolytic
FT cleavage site"
PN W09600503-A1.
PD 11-JAN-1996.
PF 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI ) MERCK & CO INC.
PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Claim 6; Page 64; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSI and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC action of prostate specific antigen (PSA), a protease with
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond

```

```

OS Homo sapiens.
PN W09600503-A1.
PD 11-JAN-1996.
PF 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI ) MERCK & CO INC.
PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Disclosure; Page 60; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSI and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC action of prostate specific antigen (PSA), a protease with
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond
SQ Sequence 17 AA;

Query Match      87.8%; Score 43; DB 17; Length 17;
Best Local Similarity 85.7%; Pred. No. 7.10e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 3 dnkisyq 9
   :|||||
QY 1 QNKISYQ 7

RESULT 4
ID R89612 standard; peptide; 7 AA.
AC R89612;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
KW gel structure; dissolution; prostate specific antigen; proteolysis;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 6..7 /note= "prostate specific antigen proteolytic
FT cleavage site"
PN W09600503-A1.
PD 11-JAN-1996.
PF 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI ) MERCK & CO INC.
PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Claim 6; Page 64; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSI and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC action of prostate specific antigen (PSA), a protease with
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond

```

CC or peptide linker, to a cytotoxic agent and used to treat
 CC prostate cancer.
 SQ Sequence 7 AA;

Query Match 85.7%; Score 42; DB 17; Length 7;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 Qy 2 NKISYQ 7

RESULT 5

ID W33449 standard; Peptide; 7 AA.
 AC W33449;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 29 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 6.7
 FT /note= "PSA specific cleavage site"
 PN W09712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; U15713.
 PR 06-OCT-1995; US-540412.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI: 97-225974/20.
 PT Oligopeptide(s) recognised and cleaved by free prostate specific antigen
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 6; Page 162; 192pp; English.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 7 AA;

Query Match 85.7%; Score 42; DB 26; Length 7;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 Qy 2 NKISYQ 7

RESULT 6

ID W33546 standard; Peptide; 8 AA.
 AC W33546;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 126 based on Semenogelin I cleavage site.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;

KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 6.7
 FT /note= "PSA specific cleavage site"
 PN W09712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; U15713.
 PR 06-OCT-1995; US-540412.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI: 97-225974/20.
 PT Oligopeptide(s) recognised and cleaved by free prostate specific
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 3; Page 160; 192pp; English.
 CC The present sequence is a novel oligopeptide designed based on the
 CC sequences surrounding the prostate specific antigen (PSA) cleavage sites
 CC of Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 8 AA;

Query Match 85.7%; Score 42; DB 26; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 Qy 2 NKISYQ 7

RESULT 7

ID W33435 standard; Peptide; 8 AA.
 AC W33435;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 15 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Modified_site 1
 FT /note= "optionally N-terminally acetylated, when
 FT C-terminally amidated"
 FT Cleavage_site 6.7
 FT /note= "PSA specific cleavage site"
 FT Modified_site 8
 FT /note= "this Ser is optionally attached via a peptide
 FT bond to the 3'-amino group on the sugar ring of
 FT doxorubicin, alternatively it is optionally
 FT amidated when the peptide is N-terminally
 FT acetylated"

PN W09712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; U15713.
 PR 06-OCT-1995; US-540412.

PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI; 97-225974/20.
 PT Oligopeptide(s) recognised and cleaved by free prostate specific
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 3 and 7 and 17; Page 160; 192pp; English.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA. In a study of cleavage affinity, the TFA salt of this peptide
 CC (with an N-terminal acetyl group and a C-terminal amide group) was
 CC digested with enzymatically active PSA. After four hours 30 per cent of
 CC the peptide salt was cleaved.
 SQ Sequence 8 AA;

Query Match 85.7%; Score 42; DB 26; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 8
 ID R89598 standard; peptide; 8 AA.
 AC R89598;
 DT 02-SEP-1996 (first entry)
 DE Prostate specific antigen, semenogelin derived, cleavage substrate.
 KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
 KW gel structure; dissolution; prostate specific antigen; proteolysis;
 KW chymotrypsin like specificity; peptide substrate; cleavage site;
 KW assay; determination; proteolytic activity; identification;
 KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cleavage_site 6..7
 FT /note= "prostate specific antigen proteolytic
 FT cleavage site"
 FT modified_site 1
 FT /note= "opt. acylated"
 FT modified_site 12
 FT /note= "opt. amidated"
 FT WO9600503-A1.
 PN 11-JAN-1996.
 PD 07-JUN-1995; U08156.
 PF 28-JUN-1994; US-267092.
 PR 15-MAR-1995; US-404833.
 PR (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI; 96-077275/08.
 PT New peptide substrates cleaved by prostate-specific antigen - also
 PT cytotoxic conjugates for treating prostate cancer, and assay for
 PT determination of PSA activity
 PS Claim 3; Page 58; 142pp; English.
 CC Human semenogelin I (hSI) is one of the major proteins, including
 CC hSI and fibronectin, in the sperm entrapping gel formed at

CC ejaculation. This gel structure undergoes dissolution via the
 CC action of prostate specific antigen (PSA), a protease with
 CC chymotrypsin like specificity, which proteolyses the above major
 CC proteins. New substrates, including the present peptide, cleaved
 CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
 CC in assays to determine the proteolytic activity of free PSA in a
 CC sample, and to identify cpds. which inhibit the proteolytic
 CC activity of PSA, they may also be conjugated, via a covalent bond
 CC or peptide linker, to a cytotoxic agent and used to treat
 CC prostate cancer. In a PSA hydrolysis assay, the percentage of
 CC the present peptide cleaved by YORK PSA after 4 hrs. was 30 %
 CC acylated and amidated.
 SQ Sequence 8 AA;

Query Match 85.7%; Score 42; DB 17; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 9
 ID R89712 standard; peptide; 8 AA.
 AC R89712;
 DT 02-SEP-1996 (first entry)
 DE Prostate specific antigen, semenogelin derived, cleavage substrate.
 KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
 KW gel structure; dissolution; prostate specific antigen; proteolysis;
 KW chymotrypsin like specificity; peptide substrate; cleavage site;
 KW assay; determination; proteolytic activity; identification;
 KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cleavage_site 6..7
 FT /note= "prostate specific antigen proteolytic
 FT cleavage site"
 FT WO9600503-A1.
 PN 11-JAN-1996.
 PD 07-JUN-1995; U08156.
 PF 28-JUN-1994; US-267092.
 PR 15-MAR-1995; US-404833.
 PR (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI; 96-077275/08.
 PT New peptide substrates cleaved by prostate-specific antigen - also
 PT cytotoxic conjugates for treating prostate cancer, and assay for
 PT determination of PSA activity
 PS Claim 3; Page 108; 142pp; English.
 CC Human semenogelin I (hSI) is one of the major proteins, including
 CC hSI and fibronectin, in the sperm entrapping gel formed at
 CC ejaculation. This gel structure undergoes dissolution via the
 CC action of prostate specific antigen (PSA), a protease with
 CC chymotrypsin like specificity, which proteolyses the above major
 CC proteins. New substrates, including the present peptide, cleaved
 CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
 CC in assays to determine the proteolytic activity of free PSA in a
 CC sample, and to identify cpds. which inhibit the proteolytic
 CC activity of PSA, they may also be conjugated, via a covalent bond
 CC or peptide linker, to a cytotoxic agent and used to treat
 CC prostate cancer.
 SQ Sequence 8 AA;

Query Match 85.7%; Score 42; DB 17; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7


```

RESULT 10
ID R89599 standard; peptide; 9 AA.
AC R89599;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; prostate;
KW gel structure; dissolution; prostate specific antigen; proteolysis;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 6..7
FT /note= "prostate specific antigen proteolytic
FT cleavage site"
PN WO9600503-A1.
PD 11-JAN-1996.
PF 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI ) MERCK & CO INC.
PI Defeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR WPI; 96-077275/08.
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Claim 3; Page 58; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSII and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC action of prostate specific antigen (PSA), a protease with
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond
CC or peptide linker, to a cytotoxic agent and used to treat
CC prostate cancer. In a PSA hydrolysis assay, the percentage of
CC the present peptide cleaved by YORK PSA after 4 hrs. was 55 %.
SQ Sequence 9 AA;

Query Match 85.7%; Score 42; DB 17; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.60e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
Qy 2 NKISYQ 7

RESULT 11
ID R89718 standard; peptide; 9 AA.
AC R89718;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; prostate;
KW gel structure; dissolution; prostate specific antigen; protease;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 7..8
FT /note= "prostate specific antigen proteolytic
FT cleavage site"
PN WO9600503-A1.
PD 11-JAN-1996.
PF 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI ) MERCK & CO INC.
PI Defeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;

```

```

DR WPI; 96-077275/08.
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Claim 4; Page 111; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSII and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC action of prostate specific antigen (PSA), a protease with
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond
CC or peptide linker, to a cytotoxic agent and used to treat
CC prostate cancer.
SQ Sequence 9 AA;

Query Match 85.7%; Score 42; DB 17; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.60e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 nkisyq 7
Qy 2 NKISYQ 7

RESULT 12
ID W33552 standard; Peptide; 9 AA.
AC W33552;
DT 27-MAR-1998 (first entry)
DE Oligopeptide 132 based on Semenogelin I cleavage site.
KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
KW detection; cytotoxic conjugate; activation; quantitative assay.
OS Synthetic.
FH Key Location/Qualifiers
FT cleavage_site 7..8
FT /note= "PSA specific cleavage site"
PN WO9712624-A1.
PD 10-APR-1997.
PF 02-OCT-1996; U15713.
PR 06-OCT-1995; US-540412.
PA (MERI ) MERCK & CO INC.
PI Defeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR WPI; 97-225974/20.
PT Oligopeptide(s) recognised and cleaved by free prostate specific
PT antigen - useful in assays for active antigen, and in
PT oligopeptide-drug conjugates for prostatic cancer treatment
PS Claim 4; Page 161; 192pp; English.
CC The present sequence is a novel oligopeptide designed based on the
CC sequences surrounding the prostate specific antigen (PSA) cleavage sites
CC of Semenogelin I, a major sperm entrapping gel protein. PSA, which has
CC chymotrypsin-like specificity is responsible for dissolution of the gel
CC structure formed at ejaculation by proteolysis of the major gel proteins
CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
CC releases progressively motile spermatozoa. PSA complexed to
CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
CC of the prostate. Prostate metastases are also known to secrete
CC immunologically reactive PSA since serum PSA is detectable at high levels
CC in prostatectomised patients showing widespread metastatic prostate
CC cancer. Cytotoxic compounds that could be activated by the proteolytic
CC activity of PSA should also be prostate cell specific as well as specific
CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
CC and this oligopeptide are useful in treatment of prostate cancer. The
CC oligopeptide can also be used in a quantitative assay for enzymatically
CC active PSA.
SQ Sequence 9 AA;

Query Match 85.7%; Score 42; DB 26; Length 9;

```

Best Local Similarity 100.0%; Pred. No. 9.50e+01; Mismatches 0; Indels 0; Gaps 0;
 Matches 6; Conservative 0;

Db 2 nkisyq 7
 |||||
 QY 2 NKISYQ 7

RESULT 13
 ID W33488 standard; Peptide; 9 AA.
 AC W33488;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 68 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 7..8 /note= "PSA specific cleavage site"
 PN WO9712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; U15713.
 PR 06-OCT-1995; US-540412.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI: 97-225974/20.
 PT Oligopeptide(s) recognised and cleaved by free prostate specific antigen.
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Disclosure; Page 102; 192pp; English.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 9 AA;

Query Match 85.7%; Score 42; DB 26; Length 9;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 nkisyq 7
 |||||
 QY 2 NKISYQ 7

RESULT 14
 ID W33448 standard; Peptide; 10 AA.
 AC W33448;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 28 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.

OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 6..7 /note= "PSA specific cleavage site"
 PN WO9712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; U15713.
 PR 06-OCT-1995; US-540412.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI: 97-225974/20.
 PT Oligopeptide(s) recognised and cleaved by free prostate specific antigen.
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 6; Page 162; 192pp; English.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 10 AA;

Query Match 85.7%; Score 42; DB 26; Length 10;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 15
 ID W33429 standard; Peptide; 10 AA.
 AC W33429;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 9 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 5..6 /note= "PSA specific cleavage site"
 FT Modified_site 10 /note= "this Thr is optionally attached via a peptide bond to the 3'-amino group on the sugar ring of doxorubicin"
 PN WO9712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; U15713.
 PR 06-OCT-1995; US-540412.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI: 97-225974/20.
 PT Oligopeptide(s) recognised and cleaved by free prostate specific antigen.
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 4 and 17; Page 161; 192pp; English.

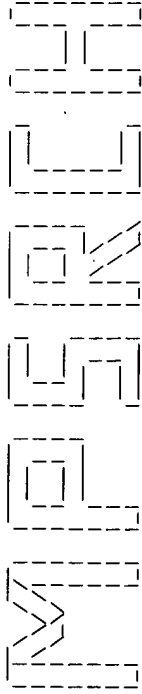
CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA. In a study of cleavage affinity, the RFA salt of this peptide
 CC was digested with enzymatically active PSA. After four hours 62 per cent
 CC of the peptide was cleaved.
 SQ Sequence 10 AA;

Query Match 85.7%; Score 42; DB 26; Length 10;
 Best Local Similarity 100.0%; Pred. No. 9.60e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

Search completed: Thu Oct 28 11:29:19 1999
 Job time : 17 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:29:36 1999; MasPar time 2.95 Seconds

Tabular output not generated. 95.143 Million cell updates/sec

Title: >US-09-081-707-11

Description: (1-7) from US09081707.pap

Perfect Score: 49

Sequence: 1 QNKISYQ 7

Scoring table: PAM 150

Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir60

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.354; Variance 20.844; scale 0.976

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	49	100.0	287	2	JE0220	2.00e+01
2	45	91.8	582	2	A43412	2.11e+00
3	45	91.8	704	2	I47228	2.11e+00
4	42	85.7	145	2	A64214	1.13e+01
5	42	85.7	462	1	WTHUB	1.13e+01
6	41	83.7	136	2	S52806	1.13e+01
7	41	83.7	136	1	ERAD29	1.93e+01
8	41	83.7	669	2	S6551	1.93e+01
9	41	83.7	825	1	GLHQ	1.93e+01
10	40	81.6	135	1	ERAD74	1.93e+01
11	40	81.6	135	2	UC4773	3.28e+01
12	40	81.6	135	2	C37476	3.28e+01
13	40	81.6	142	2	S56414	3.28e+01
14	40	81.6	326	2	H71497	3.28e+01
15	40	81.6	344	2	A42359	3.28e+01
16	40	81.6	553	2	I49834	3.28e+01
17	40	81.6	709	2	E54213	3.28e+01
18	40	81.6	719	1	MXR3D	3.28e+01
19	40	81.6	967	2	A64710	3.28e+01
20	40	81.6	969	2	JF1810	3.28e+01
21	39	79.6	129	2	JG0019	5.51e+01
22	39	79.6	221	2	S41387	5.51e+01
23	39	79.6	248	2	B71664	5.51e+01

24 39 79.6 293 2 C71703 hypothetical protein 5.51e+01
25 39 79.6 384 2 A32885 xylose operon repress 5.51e+01
26 39 79.6 557 2 S58154 glucose-6-phosphate i 5.51e+01
27 39 79.6 1021 2 E64576 hypothetical protein 5.51e+01
28 39 79.6 1062 2 B26330 hypothetical protein 5.51e+01
29 39 79.6 1089 2 S48244 NMD2 protein - yeast 5.51e+01
30 39 79.6 1234 2 B36186 I factor 2 (transposo 5.51e+01
31 39 79.6 2178 2 S55805 alpha-toxin - Clostri 5.51e+01
32 39 79.6 2567 2 A49551 filamin, Mueller cell 5.51e+01
33 39 79.6 6669 2 S55024 nebulin, skeletal mus 5.51e+01
34 38 77.6 149 2 S75155 hypothetical protein 9.16e+01
35 38 77.6 312 1 S28006 RNA-directed DNA poly 9.16e+01
36 38 77.6 337 2 S36824 beta-galactoside alph 9.16e+01
37 38 77.6 380 2 H70069 capsular polyglutamat 9.16e+01
38 38 77.6 385 2 T01332 chalcone synthase hom 9.16e+01
39 38 77.6 430 2 A34252 long-chain-acyl-coA d 9.16e+01
40 38 77.6 431 2 A56822 synaptonemal complex 9.16e+01
41 38 77.6 439 2 JC4036 phosphopyruvate hydra 9.16e+01
42 38 77.6 490 2 S70352 protein kinase clk3 (9.16e+01
43 38 77.6 612 2 S73611 glucose inhibited div 9.16e+01
44 38 77.6 1021 2 S64506 protein kinase BUB1 (9.16e+01
45 38 77.6 1299 2 T00261 hypothetical protein 9.16e+01

ALIGNMENTS

RESULT 1
ENTRY JE0220 #type complete
TITLE 28k surface antigen I - Ehrlichia canis
ORGANISM #formal_name Ehrlichia canis
DATE 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999

ACCESSIONS JE0220
REFERENCE JE0216
#authors Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
#journal Biochem. Biophys. Res. Commun. (1998) 247:636-643
#title Molecular characterization of a 28kDa surface antigen gene family of the tribe Ehrlichiae.
#cross-references MUID:98321180
#accession JE0220
##molecule_type DNA
##residues 1-287 #label RED
##cross-references GB:AF062762
SUMMARY #length 287 #molecular-weight 32014 #checksum 463

Query Match 100.0%; Score 49; DB 2; Length 287;
Best Local Similarity 100.0%; Pred. No. 2.00e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 214 QNKISYQ 220
| | | | |
QY 1 QNKISYQ 7

RESULT 2
ENTRY A43412 #type complete
TITLE semenogelin II precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-May-1998

ACCESSIONS A43412; B31489; A45295; S29156; S68765; S68762
REFERENCE A43412
#authors Ulvsback, M.; Lazure, C.; Lilja, H.; Spurr, N.K.; Rao, V.V.; Loffler, C.; Hansmann, I.; Lundwall, A.
#journal J. Biol. Chem. (1992) 267:18080-18084
#title Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20.

#cross-references MUID:92388176
#accession A43412
##molecule_type DNA
##residues 1-582 #label ULV

```

##cross-references GB:M81651; NID:g307417; PID:g307418
##note sequence extracted from NCBI backbone (NCBIN:112887, NCBIP:112889)

REFERENCE
#authors A31489
#journal Lilja, H.; Abrahamsson, P.A.; Lundwall, A.
#title J. Biol. Chem. (1989) 264:1894-1900
#title Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa.

#cross-references MUID:89109215
#accession B31489
##status nucleic acid sequence not shown
##molecule_type mRNA
##residues 214-278, 'Y', 280-281 ##label LIL

REFERENCE
#authors A45295
#journal Lilja, H.; Lundwall, A.
#title Proc. Natl. Acad. Sci. U.S.A. (1992) 89:4559-4563
#title Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein.

#cross-references MUID:92262479
#accession A45295
##molecule_type mRNA
##residues 3-582 ##label LI2
##cross-references GB:M81652

REFERENCE
#authors S29155
#journal Schneider, K.; Kausler, W.; Tripiet, D.; Jouvenal, K.; Spittler, G.
#title Biol. Chem. Hoppe-Seyler (1989) 370:353-356
#title Isolation and structure determination of two peptides occurring in human seminal plasma.

#accession S29156
##molecule_type protein
##residues 390, 'E', 392-396, 'E', 398-400, 'EW', 403-405, 'C', 407, 'DE' ##label SCH
##note this report is of a secondary sequence determined simultaneously with the sequence with accession number S29155 (see entry WTU8); as a secondary sequence, it should be considered less than fully reliable

REFERENCE
#authors S68765
#journal Kise, H.; Nishioka, J.; Kawamura, J.; Suzuki, K.
#title Eur. J. Biochem. (1996) 238:88-96
#title Characterization of semenogelin II and its molecular interaction with prostate-specific antigen and protein C inhibitor.

#cross-references MUID:96248425
#accession S68765
##molecule_type protein
##residues 105-107, 'K', 109-111; 113-122; 260-269; 280-283 ##label KIS

REFERENCE
#authors S68761
#journal Malm, J.; Hellman, J.; Magnusson, H.; Laurell, C.B.; Lilja, H.
#title Eur. J. Biochem. (1996) 238:48-53
#title Isolation and characterization of the major gel proteins in human semen, semenogelin I and semenogelin II.

#accession S68762
##molecule_type protein
##residues 420-421, 'G', 423-423 ##label MAL

GENETICS
#gene GDB:SEMG2
#map_position 20q12-20q13.1
#introns 26/1
#CLASSIFICATION superfamily semenogelin
#KEYWORDS duplication; glycoprotein; semen; seminal vesicle; tandem repeat

FEATURE
1-23 #domain signal sequence #status predicted #label SIG\
24-582 #product semenogelin II #status predicted #label MAT\
272 #binding_site carbohydrate (Asn) (covalent) #status Predicted

SUMMARY
#length 582 #molecular-weight 65444 #checksum 9704
91.8%; Score 45; DB 2; Length 582;

Query Match

```

```

Best Local Similarity 85.7%; Pred. No. 2.11e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 343 ENKISYQ 349
Qy 1 QNKISYQ 7
:|||||
:|||||

RESULT 3
ENTRY #type complete
TITLE carbonic anhydrase II inhibitor (transferrin homolog)
ORGANISM precursor - pig
DATE #formal_name Sus scrofa domestica #common_name domestic pig
21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 15-Jan-1999
ACCESSIONS I47228
REFERENCE I47228
#authors Roush, E.D.; Fierke, C.A.
#journal Biochemistry (1992) 31:12536-12542
#title Purification and characterization of a carbonic anhydrase II inhibitor from porcine plasma.

#cross-references MUID:93099129
#accession I47228
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-704 ##label ROU
##cross-references EMBL:U36916; NID:g1016329; PID:g1016330

GENETICS
#gene PICA
#CLASSIFICATION superfamily transferrin; transferrin repeat homology
#KEYWORDS duplication
#FEATURE
20-350 #domain transferrin repeat homology #label TRHI
SUMMARY #length 704 #molecular-weight 77634 #checksum 8035

Query Match
Best Local Similarity 71.4%; Score 45; DB 2; Length 704;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 667 QDKITYQ 673
Qy 1 QNKISYQ 7
:|||||
:|||||

RESULT 4
ENTRY #type complete
TITLE hypothetical protein homolog MG127 - Mycoplasma genitalium (SGC3)
ORGANISM #formal_name Mycoplasma genitalium
DATE 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 09-Apr-1998
ACCESSIONS A64214
REFERENCE A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession A64214
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-145 ##label TIGR
##cross-references GB:U36911; GB:L43967; NID:g1045794; PID:g1045807; TIGR:MG127
##experimental_source strain G-37

GENETICS

```

```

#genetic_code SGC3
CLASSIFICATION #superfamily hypothetical protein yjbd
SUMMARY #length 145 #molecular-weight 16565 #checksum 2477

Query Match 85.7%; Score 42; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 1.13e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 QNKISY 39
| | | | |
QY 1 QNKISY 6

RESULT 5
ENTRY #type complete
TITLE semenogelin I precursor - human
CONTAINS seminal basic protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 04-Dec-1986 #sequence_revision 03-Oct-1995 #text_change
26-Feb-1999
ACCESSIONS B43412; A31489; A91335; S29155; A43500; A91320; S29380;
S68761; A03254
A43412
REFERENCE
#authors Ulysback, M.; Lazure, C.; Lilja, H.; Spurr, N.K.; Rao, V.V.;
Loffler, C.; Hansmann, I.; Lundwall, A.
#journal J. Biol. Chem. (1992) 267:18080-18084
#title Gene structure of semenogelin I and II. The predominant
proteins in human semen are encoded by two homologous genes
on chromosome 20.
#cross-references MUID:92388176
#accession B43412
#molecule_type DNA
#residues 1-462 #label ULV
#cross-references GB:M81650; NID:9307416; PID:9487420
REFERENCE
#authors Lilja, H.; Abrahamsson, P.A.; Lundwall, A.
#journal J. Biol. Chem. (1989) 264:1894-1900
#title Semenogelin, the predominant protein in human semen. Primary
structure and identification of closely related proteins in
the male accessory sex glands and on the spermatozoa.
#cross-references MUID:89109215
#accession A31489
#molecule_type mRNA
#residues 1-78,'T',80-422,'K',424-462 #label LIL
#cross-references GB:J04440
#note Ser-79 was also found
REFERENCE
#authors A91335
#journal Lilja, H.; Jeppsson, J.O.
#journal FEBS Lett. (1985) 182:181-184
#title Amino acid sequence of the predominant basic protein in human
seminal plasma.
#cross-references MUID:85127550
#accession A91335
#molecule_type protein
#residues 108-159 #label LI2
#note this sequence represents a naturally occurring fragment
from proteolytic cleavage of semenogelin during
liquefaction of semen
REFERENCE
#authors S29155
#journal Schneider, K.; Kausler, W.; Tripiel, D.; Jouvenal, K.;
Spiteller, G.
#journal Biol. Chem. Hoppe-Seyler (1989) 370:353-356
#title Isolation and structure determination of two peptides
occurring in human seminal plasma.
#accession S29155
#molecule_type protein
#residues 316-320,'L',322-344 #label SCH
#note this sequence represents the amino end of a naturally
occurring fragment from proteolytic cleavage of
semenogelin during liquefaction of semen
REFERENCE
#authors A43500
#journal Ramasharma, K.; Sairam, M.R.; Seidah, N.G.; Chretien, M.;
Manjunath, P.; Schiller, P.W.; Yamashiro, D.; Li, C.H.

```

```

#journal Science (1984) 223:1199-1202
#title Isolation, structure, and synthesis of a human seminal plasma
peptide with inhibin-like activity.
#cross-references MUID:84146751
#accession A43500
#molecule_type protein
#residues 108-138 #label RAM
#note this sequence represents a naturally occurring fragment
from proteolytic cleavage of semenogelin during
liquefaction of semen
REFERENCE
#authors A91320
#journal Seidah, N.G.; Ramasharma, K.; Sairam, M.R.; Chretien, M.
#journal FEBS Lett. (1984) 167:98-102
#title Partial amino acid sequence of a human seminal plasma peptide
with inhibin-like activity.
#cross-references MUID:84132557
#accession A91320
#molecule_type protein
#residues 108-138 #label SEI
#note this sequence represents a naturally occurring fragment
from proteolytic cleavage of semenogelin during
liquefaction of semen
REFERENCE
#authors S29380
#journal Khan, Z.; Smyth, D.G.
#journal Eur. J. Biochem. (1993) 212:35-40
#title Isolation and identification of N-terminally extended forms
of 5-oxoprollylglutamylprolinamide (Glp-Glu-Pro-NH(2)), a
thyrotropin-releasing-hormone(TRH)-like peptide present in
human semen.
#cross-references MUID:93185635
#accession S29380
#molecule_type protein
#residues 373-397 #label KHA
#note the authors' suggestion that this peptide is amidated is
consistent with radioimmunoassay results but has not
been proven
#note the amidated tripeptide Glp-Glu-Pro-NH2 (where Glp is
pyroglutamic acid) is present in human semen; although
the sequence reported here is consistent with an
amino-terminally extended form derived from
semenogelin, the sequence is followed in semenogelin
by Trp, which (unlike Glp) probably cannot become the
source of an amide moiety; the authors conclude the
peptide must be derived from a closely related protein
REFERENCE
#authors S68761
#journal Malm, J.; Hellman, J.; Magnusson, H.; Laurell, C.B.; Lilja,
H.
#journal Eur. J. Biochem. (1996) 238:48-53
#title Isolation and characterization of the major gel proteins in
human semen, semenogelin I and semenogelin II.
#accession S68761
#molecule_type protein
#residues 49-50,'G',52-53 #label MAL
#note This abundant protein from seminal vesicle secretions maintains a
gel-like environment for the sperm cells. At ejaculation,
kallikrein-like enzymes in prostatic secretions cleave this
protein, resulting in liquefaction of the seminal gel and
allowing increased sperm motility.
GENETICS
#gene GDB:SEMG1
#cross-references GDB:128167; OMIM:182140
#map_position 20q12-20q13.1
#introns 26/1
CLASSIFICATION #superfamily semenogelin
KEYWORDS duplication; glycoprotein; semen; seminal vesicle
FEATURE
1-24 #domain signal sequence #status predicted #label SIG\
108-138 #product seminal basic protein #status experimental
#label MAL\
174-215 #region semenogelin short repeat 1\
234-275 #region semenogelin short repeat 2\
282-339 #region semenogelin long repeat 1\
342-399 #region semenogelin long repeat 2\

```

```
414-455      #region semenogelin short repeat 3\
141          #binding_site carbohydrate (Asn) (covalent) #status
239          #disulfide_bonds interchain #status experimental
SUMMARY      #length 462 #molecular-weight 52117 #checksum 9901

Query Match      85.7%; Score 42; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.13e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 284 NKISYQ 289
QY 2 NKISYQ 7

RESULT 6
ENTRY S52806 #type complete
TITLE hypothetical protein, 15.3K - human adenovirus 7
ORGANISM #formal_name Mastadenovirus h7 #common_name human adenovirus
DATE 19-May-1995 #sequence_revision 21-Jul-1995 #text_change
08-Sep-1997
ACCESSIONS S52806
REFERENCE S52798
#authors Kajon, A.E.; Wadell, G.
#submission submitted to the EMBL Data Library, April 1995
#description Sequence analysis of the E3 region and fiber gene of human
adenovirus 7h.
#accession S52806
#molecule_type DNA
#residues
##cross-references EMBL:Z48954; NID:g762955; PID:g762964
##experimental_source strain 87-922
CLASSIFICATION #superfamily adenovirus early E3B 14K protein
SUMMARY #length 136 #molecular-weight 15275 #checksum 7906

Query Match      83.7%; Score 41; DB 2; Length 136;
Best Local Similarity 71.4%; Pred. No. 1.93e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 62 QKISYE 68
QY 1 QNKISYQ 7

RESULT 7
ENTRY ERAD29 #type complete
TITLE early E3B 14K protein - human adenovirus 3
ALTERNATE_NAMES early E3B 15.3K protein
ORGANISM #formal_name Mastadenovirus h3 #common_name human adenovirus
DATE 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
07-Nov-1997
ACCESSIONS A27623
REFERENCE A91566
#authors Signaes, C.; Akusjaervi, G.; Pettersson, U.
#journal Gene (1986) 50:173-184
#title Region E3 of human adenoviruses; differences between the
oncogenic adenovirus-3 and the non-oncogenic adenovirus-2.
#cross-references MUID:87219876
#accession A27623
#molecule_type DNA
#residues 1-136 #label SIG
##cross-references GB:M15952; NID:g209901; PID:g209910
CLASSIFICATION #superfamily adenovirus early E3B 14K protein
KEYWORDS early protein
SUMMARY #length 136 #molecular-weight 15266 #checksum 8383

Query Match      83.7%; Score 41; DB 1; Length 136;
Best Local Similarity 71.4%; Pred. No. 1.93e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 62 QKISYE 68
```

```
QY 1 QNKISYQ 7

RESULT 8
ENTRY S65551 #type fragment
TITLE factor H - bovine (fragment)
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
16-Dec-1998
ACCESSIONS S65551
REFERENCE S65551
#authors Soames, C.J.; Day, A.J.; Sim, R.B.
#journal Biochem. J. (1996) 315:523-531
#title Prediction from sequence comparisons of residues of factor H
involved in the interaction with complement component C3b.
#accession S65551
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-569 #label SOA
CLASSIFICATION #superfamily complement factor H; complement factor H repeat
homology
FEATURE 296-349 #domain complement factor H repeat homology #label FH02\
355-412 #domain complement factor H repeat homology #label FH08\
416-471 #domain complement factor H repeat homology #label FH09\
538-592 #domain complement factor H repeat homology #label FH11\
599-651 #domain complement factor H repeat homology #label FH07\
SUMMARY #length 669 #checksum 8335

Query Match      83.7%; Score 41; DB 2; Length 669;
Best Local Similarity 57.1%; Pred. No. 1.93e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 196 QDKITYE 202
QY 1 QNKISYQ 7

RESULT 9
ENTRY GLHQ #type complete
TITLE beta-glucosidase (EC 3.2.1.21) precursor - yeast (Pichia
anomala)
ALTERNATE_NAMES beta-D-glucoside glucosylhydrolase; cellobiase; gentiobiase
ORGANISM #formal_name Pichia anomala, Candida pelliculosa
DATE 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change
01-Nov-1996
ACCESSIONS B23783
REFERENCE A93583
#authors Kohchi, C.; Toh-e, A.
#journal Nucleic Acids Res. (1985) 13:6273-6282
#title Nucleotide sequence of Candida pelliculosa beta-glucosidase
gene.
#cross-references MUID:86016087
#accession B23783
#molecule_type DNA
#residues 1-825 #label KOH
##experimental_source var acetabacterius
##note the authors translated the codon TGT for residue 320 as
Gly, TAT for residue 356 as Thr, GAA for residue 375
as Gln, AAT for residue 418 as Asp, and GAG for
residue 487 as Gly
COMMENT This enzyme is secreted into the periplasmic space, which suggests
the presence of a signal sequence.
CLASSIFICATION #superfamily beta-glucosidase
KEYWORDS glycoprotein; glycosidase; hydrolase; polysaccharide
degradation
FEATURE 1-17 #domain signal sequence #status predicted #label SIG\
18-825 #product beta-glucosidase #status predicted #label MAT\
21,34,74,97,230,
271,328,335,537,
550,578,667,690,
```



```

718,733,761      #binding-site carbohydrate (Asn) (covalent) #status
                  predicted
SUMMARY          #length 825 #molecular-weight 89560 #checksum 8818

Query Match      83.7%; Score 41; DB 1; Length 825;
Best Local Similarity 71.4%; Pred. No. 1.93e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 486 QEKISYE 492
   1 QNKISYQ 7

RESULT 10
ENTRY   ERAD74      #type complete
TITLE   early E3B 15.3k protein - human adenovirus 7 (strain Gomen)
ORGANISM #formal_name Mastadenovirus h7 #common_name human adenovirus
#note    host Homo sapiens (man)
DATE     31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
ACCESSION E31830
REFERENCE A94386
#authors  Hong, J.S.; Mullis, K.G.; Engler, J.A.
#journal  Virology (1988) 167:545-553
#title    Characterization of the early region 3 and fiber genes of
          Ad7.
#cross-references MUID:89073758
#accession E31830
#molecule_type DNA
##residues 1-135 ##label HON
##cross-references GB:M23696; NID:g341012; PID:g576460
CLASSIFICATION #superfamily adenovirus early E3B 14K protein
KEYWORDS early protein
SUMMARY #length 135 #molecular-weight 15363 #checksum 8529

Query Match      81.6%; Score 40; DB 1; Length 135;
Best Local Similarity 71.4%; Pred. No. 3.28e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 61 QAKISYE 67
   1 QNKISYQ 7

RESULT 11
ENTRY   JC4773      #type complete
TITLE   E3 gene encoding hypothetical 15.3k protein - human
          adenovirus 35
ORGANISM #formal_name Mastadenovirus h35 #common_name human adenovirus
          35
DATE     10-May-1996 #sequence_revision 16-Aug-1996 #text_change
ACCESSION JC4773
REFERENCE JC4773
#authors  Basler, C.F.; Droguett, G.; Horwitz, M.S.
#journal  Gene (1996) 170:249-254
#title    Sequence of the immunoregulatory early region 3 and flanking
          sequences of adenovirus type 35.
#cross-references MUID:96233144
#accession JC4773
#molecule_type DNA
##residues 1-135 ##label BAS
##cross-references GB:U32664; NID:g984529; PID:g984538
COMMENT      This protein inhibits TNF-alpha cytolysis.
GENETICS
#gene      E3
CLASSIFICATION #superfamily adenovirus early E3B 14K protein
SUMMARY #length 135 #molecular-weight 15321 #checksum 7832

Query Match      81.6%; Score 40; DB 2; Length 135;
Best Local Similarity 71.4%; Pred. No. 3.28e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 61 QAKISYE 67
   1 QNKISYQ 7

RESULT 12
ENTRY   C37476      #type complete
TITLE   hypothetical protein, 15.3K - human adenovirus 11
ORGANISM #formal_name Mastadenovirus h11 #common_name human adenovirus
          11
DATE     21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
          03-May-1996
ACCESSION C37476
REFERENCE A37476
#authors  Mei, Y.F.; Wadell, G.
#journal  Virology (1993) 194:453-462
#title    Hemagglutination properties and nucleotide sequence analysis
          of the fiber gene of adenovirus genome types 11p and 11a.
#cross-references MUID:93276532
#contents  Adlip, Slobitski
#accession C37476
#status    preliminary
#molecule_type nucleic acid
##residues 1-135 ##label MEI
##note     sequence extracted from NCBI backbone (NCBIN:132883,
          NCBI:P:132886)
CLASSIFICATION #superfamily adenovirus early E3B 14K protein
SUMMARY #length 135 #molecular-weight 15293 #checksum 7034

Query Match      81.6%; Score 40; DB 2; Length 135;
Best Local Similarity 71.4%; Pred. No. 3.28e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 61 QAKISYE 67
   1 QNKISYQ 7

RESULT 13
ENTRY   S56414      #type complete
TITLE   hypothetical 16K protein (aidB-rpsF intergenic region) -
          Escherichia coli
ORGANISM #formal_name Escherichia coli
          14-Nov-1997
DATE     28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
ACCESSION S56414; H65229
REFERENCE S56314
#authors  Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.;
          Blattner, F.R.
#journal  Nucleic Acids Res. (1995) 23:2105-2119
#title    Analysis of the Escherichia coli genome VI: DNA sequence of
          the region from 92.8 through 100 minutes.
#cross-references MUID:95334362
#accession S56414
#status    preliminary; nucleic acid sequence not shown;
          translation not shown
#molecule_type DNA
##residues 1-142 ##label BUR
##cross-references EMBL:U14003; NID:g1263172; PID:g537030
##note     the nucleotide sequence was submitted to the EMBL Data
          Library, August 1994
REFERENCE A64720
#authors  Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
          Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
          Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
          Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
          Y.
#journal  Science (1997) 277:1453-1462
#title    The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession H65229

```

```
##status      preliminary; nucleic acid sequence not shown;
##molecule_type DNA
##residues    1-142 #label BLAT
##cross-references GB:AE000491; GB:U00096; NID:g2367357; PID:g1790633;
                UWGP:b4189
##experimental_source strain K-12, substrain MGL655
GENETICS
#gene
SUMMARY
#length 142 #molecular-weight 16034 #checksum 6007
Query Match      81.6%; Score 40; DB 2; Length 142;
Best Local Similarity 57.1%; Pred. No. 3.28e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 10 HSK1AYQ 16
   :|||:
QY 1 QNKISYQ 7

RESULT 14
ENTRY H71497 #type complete
TITLE probable yop translocation J - Chlamydia trachomatis
ORGANISM (serotype D, strain UW3/Cx)
DATE #formal_name Chlamydia trachomatis
13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
21-Nov-1998
ACCESSIONS H71497
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession H71497
##status preliminary
##molecule_type DNA
##residues 1-326 #label ARN
##cross-references GB:AE001327; GB:AE001273; NID:g3328999; PID:g3329000
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene
SUMMARY
#length 326 #molecular-weight 35553 #checksum 5831
Query Match      81.6%; Score 40; DB 2; Length 326;
Best Local Similarity 71.4%; Pred. No. 3.28e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 113 QEKIRYQ 119
   :|||:
QY 1 QNKISYQ 7

RESULT 15
ENTRY A42359 #type complete
TITLE fimbrial adhesin Fl17-G precursor - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change
24-Sep-1998
ACCESSIONS A42359
REFERENCE A42359
#authors Lintermans, P.F.; Bertels, A.; Schlicker, C.; Deboeck, F.;
Charlier, G.; Pohl, P.; Norgren, M.; Normark, S.; Van
Montagu, M.; De Greve, H.
#journal J. Bacteriol. (1991) 173:3366-3373
#title Identification, characterization, and nucleotide sequence of
the Fl17-G gene, which determines receptor binding of
Escherichia coli Fl17 fimbriae.
#cross-references MUID:91258316
#accession A42359
##status preliminary
```

```
##molecule_type DNA
##residues 1-344 #label LIN
##cross-references GB:AF022140; GB:M62503; NID:g2465781; PID:g2465785
SUMMARY #length 344 #molecular-weight 36555 #checksum 2641
Query Match      81.6%; Score 40; DB 2; Length 344;
Best Local Similarity 71.4%; Pred. No. 3.28e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 57 RNKIGYQ 63
   :|||:
QY 1 QNKISYQ 7

Search completed: Thu Oct 28 11:29:51 1999
Job time : 15 secs.
```

W E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:30:07 1999; MasPar time 2.09 Seconds
Tabular output not generated. 94.559 Million cell updates/sec

Title: >US-09-081-707-11
Description: (1-7) from US09081707.pep
Perfect Score: 49
Sequence: 1 QNKISYQ 7

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 20.894; Variance 18.348; scale 1.139

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	91.8	582	1 SEM2_HUMAN	SEMNOCHELIN II PRECURS	5.26e-01
2	45	91.8	706	1 SEM2_WACMU	SEMNOCHELIN II PRECURS	5.26e-01
3	42	85.7	145	1 Y127_MYCGE	HYPOTHETICAL PROTEIN M	3.50e+00
4	42	85.7	462	1 SEM1_HUMAN	SEMNOCHELIN I PROTEIN	3.50e+00
5	41	83.7	136	1 E314_ADE03	EARLY E3 15.3 KD PROTE	6.44e+00
6	41	83.7	825	1 BGLS_HANAN	BETA-GLUCOSIDASE PRECU	6.44e+00
7	40	81.6	107	1 DGLA_STAXY	HYPOTHETICAL 12.1 KD P	1.17e+01
8	40	81.6	135	1 E314_ADE07	EARLY E3 15.3 KD PROTE	1.17e+01
9	40	81.6	142	1 YJFO_ECOLI	HYPOTHETICAL 16.0 KD P	1.17e+01
10	40	81.6	553	1 VH65_NPVAC	EARLY 65 KD PROTEIN.	1.17e+01
11	40	81.6	709	1 TOP1_MYCGE	DNA TOPOISOMERASE I (E	1.17e+01
12	40	81.6	719	1 VM3_REOVD	MAJOR NONSTRUCTURAL PR	1.17e+01
13	39	79.6	129	1 FLGB_BACSU	FLAGELLAR BASAL-BODY R	2.10e+01
14	39	79.6	384	1 XYIR_BACSU	XYLOSE REPRESSOR.	2.10e+01
15	39	79.6	487	1 ACH6_CAEEL	ACETYLCHOLINE RECEPT	2.10e+01
16	39	79.6	557	1 G6P1_ACICA	GLUCOSE-6-PHOSPHATE IS	2.10e+01
17	39	79.6	622	1 PKNB_MYCLE	PUTATIVE SERINE/THREON	2.10e+01
18	39	79.6	1089	1 NMD2_YEAST	NONSENSE-MEDIATED MRNA	2.10e+01
19	38	77.6	221	1 YMW7_YEAST	HYPOTHETICAL 25.6 KD P	3.71e+01
20	38	77.6	263	1 YHCK_ECOLI	HYPOTHETICAL TRANSCRIP	3.71e+01
21	38	77.6	295	1 Y336_MYCGE	HYPOTHETICAL PROTEIN M	3.71e+01
22	38	77.6	319	1 Y028_METJA	HYPOTHETICAL PROTEIN M	3.71e+01
23	38	77.6	337	1 CAG4_MOUSE	CMP-N-ACETYLNEURAMINAT	3.71e+01

ALIGNMENTS

RESULT 1

ID SEM2_HUMAN STANDARD: PRT; 582 AA.
AC Q02383;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SEMNOGELIN II PRECURSOR (SGII).
GN SEMG2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEMINAL VESICLE;
RX MEDLINE; 92262479.
RA LUNDWALL A., LILJA H.;
RT "Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein."
RL PROC. NATL. ACAD. SCI. U.S.A. 89:4559-4563(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92388176.
RA ULVSBAECK M., LAZURE C., LILJA H., SPURR N.K., RAO V.V., LOEFFLER C., HANSMANN I., LUNDWALL A.;
RT "Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20.";
RL J. BIOL. CHEM. 267:18080-18084(1992).
CC -!- FUNCTION: THREE SEMNOGELIN PROTEINS ARE FOUND IN HUMAN SEMEN, OF WHICH THE 52 KD SEMNOGELIN I IS THE MOST ABUNDANT. THEY PARTICIPATE IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS OF SEMNOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE PROTEASE.
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEXES OF SEMNOGELIN I AND THE OTHER SEMNOGELIN PROTEINS, THE 71- AND 76-KD POLYPEPTIDES.
CC -!- TISSUE SPECIFICITY: SEMINAL VESICLES, AND TO A MUCH LESSER EXTENT, EPIDIDYMIS.
CC -!- PFM: SEMNOGELIN II IS THOUGHT TO FORM BOTH THE 71-KD POLYPEPTIDE AND, IN ITS GLYCOSYLATED FORM, THE 76-KD POLYPEPTIDE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

24	38	77.6	426	1 C15Y_HELPY	CITRATE SYNTHASE (EC 4	3.71e+01
25	38	77.6	430	1 ACCL_RAT	ACYL-COA DEHYDROGENASE	3.71e+01
26	38	77.6	431	1 SC65_RAT	SYNAPTONEMAL COMPLEX P	3.71e+01
27	38	77.6	439	1 ENOL_SCHPO	ENOLASE (EC 4.2.1.11)	3.71e+01
28	38	77.6	490	1 CLK3_RAT	PROTEIN KINASE CLK3 (E	3.71e+01
29	38	77.6	490	1 CLK3_MOUSE	PROTEIN KINASE CLK3 (E	3.71e+01
30	38	77.6	574	1 YYP3_CAEEL	PUTATIVE SERINE CARBOX	3.71e+01
31	38	77.6	612	1 GIDA_MYCPN	GLUCOSE INHIBITED DIVI	3.71e+01
32	38	77.6	658	1 REP_BUCAP	ATP-DEPENDENT DNA HELI	3.71e+01
33	38	77.6	666	1 Y32E_MYCPN	HYPOTHETICAL PROTEIN M	3.71e+01
34	38	77.6	762	1 E13B_TRIHA	GLUCAN ENDO-1,3-BETA-G	3.71e+01
35	38	77.6	1021	1 BUB1_YEAST	CHECKPOINT SERINE/THRE	3.71e+01
36	37	75.5	225	1 YVQH_BACSU	HYPOTHETICAL 25.7 KD P	6.47e+01
37	37	75.5	334	1 PSTS_HAEIN	PHOSPHATE-BINDING PERI	6.47e+01
38	37	75.5	341	1 RIR2_HELPY	RIBONUCLEOSIDE-DIPHOSP	6.47e+01
39	37	75.5	346	1 PSTS_ECOLI	PHOSPHATE-BINDING PERI	6.47e+01
40	37	75.5	382	1 YLJ0_CAEEL	HYPOTHETICAL 44.3 KD P	6.47e+01
41	37	75.5	445	1 WAPA_STRMU	WALL-ASSOCIATED PROTEI	6.47e+01
42	37	75.5	510	1 YFK9_YEAST	HYPOTHETICAL 59.4 KD P	6.47e+01
43	37	75.5	673	1 REP_ECOLI	ATP-DEPENDENT DNA HELI	6.47e+01
44	37	75.5	891	1 SECA_PAVLU	PREPROTEIN TRANSLOCASE	6.47e+01
45	37	75.5	995	1 DP01_KLUOA	DNA POLYMERASE (EC 2.7	6.47e+01

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL: M81651; G307418; -
DR EMBL: M81652; G338239; -
DR EMBL: 247556; E82662; -
DR PIR: A43412; A43412.
DR MIM: I82141; -
KW SEMEN: SEMINAL VESICLE; REPEAT; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 23
FT CHAIN 24 582 SEMENOGELIN II.
FT DOMAIN 70 559 REPEAT-RICH REGION.
FT REPEAT 70 129 3-1.
FT REPEAT 141 200 2-1.
FT REPEAT 201 260 2-2.
FT DOMAIN 261 500 4 X 60 AA TANDEM REPEATS, TYPE I.
FT REPEAT 501 559 3-2.
FT CARBOHYD 272 272 PROBABLE.
SQ SEQUENCE 582 AA; 65445 MW; DD20304E CRC32;

Query Match 91.8%; Score 45; DB 1; Length 582;

Best Local Similarity 85.7%; Pred. No. 5.26e-01;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 343 ENKISYQ 349

:|||||

QY 1 QNKISYQ 7

RESULT 2

ID SEM2_MACMU STANDARD; PRT; 706 AA.

AC Q95196;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE SEMENOGELIN II PRECURSOR (SGII).

GN SEMG2.

OS MACACA MULATTA (RHESUS MACAQUE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 97274635.

RA ULVSBACK M., LUNDWALL A.;

RT "Cloning of the semenogelin II gene of the rhesus monkey.

RT Duplications of 360 bp extend the coding region in man, rhesus monkey

RT and baboon."

RL EUR. J. BIOCHEM. 245:25-31(1997).

CC -!- FUNCTION: PARTICIPATE IN THE FORMATION OF A GEL MATRIX ENTRAPPING

CC THE ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X92589; E208370; -

KW SEMEN: SEMINAL VESICLE; REPEAT; SIGNAL; GLYCOPROTEIN.

FT SIGNAL 1 23 POTENTIAL.

FT CHAIN 24 706 SEMENOGELIN II.

SQ SEQUENCE 706 AA; 79875 MW; 9D5897A8 CRC32;

Query Match 91.8%; Score 45; DB 1; Length 706;

Best Local Similarity 85.7%; Pred. No. 5.26e-01;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 347 ENKISYQ 353

:|||||

QY 1 QNKISYQ 7

RESULT 3
ID Y127_MYCGE STANDARD; PRT; 145 AA.
AC P47373;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MG127.
GN MG127.
OS MYCOPLASMA GENITALIUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE: 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium";
RL SCIENCE 270:397-403(1995).
CC -!- SIMILARITY: TO B.SUBTILIS YQGZ.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL: U39691; G1045807; -
DR TIGR: MG127; -
KW HYPOTHETICAL PROTEIN

SQ SEQUENCE 145 AA; 16565 MW; D203137F CRC32;

Query Match 85.7%; Score 42; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 3.50e+00;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 QNKISY 39

:|||||

QY 1 QNKISY 6

RESULT 4

ID SEM1_HUMAN STANDARD; PRT; 462 AA.

AC P04279;

DT 20-MAR-1987 (REL. 04, CREATED)

DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE SEMENOGELIN I PROTEIN PRECURSOR (SGI) [CONTAINS: SEMINAL BASIC

DE PROTEIN; ALPHA-INHIBIN-92; ALPHA-INHIBIN-31].

GN SEMGI OR SEMG.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 89109215.

RA LILJA H., ABRAHAMSSON P.-A., LUNDWALL A.;

RT "Semenogelin, the predominant protein in human semen. Primary
RT structure and identification of closely related proteins in the male
RT accessory sex glands and on the spermatozoa.";

RL J. BIOL. CHEM. 264:1894-1900(1989).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=BLOOD;

RX MEDLINE: 92388176.

RA ULVSBACK M., LAZURE C., LILJA H., SPURR N.K., RAO V.V., LOEFFLER C.,

RA HANSMANN I., LUNDWALL A.;
PT "Gene structure of semenogelin I and II. The predominant proteins in
RL human semen are encoded by two homologous genes on chromosome 20.";
RL J. BIOL. CHEM. 267:18080-18084(1992).
[3]
RN SEQUENCE OF 108-159.
RX MEDLINE; 85127550.
RA LILJA H., JEPSSON J.-O.;
RT "Amino acid sequence of the predominant basic protein in human
RL seminal plasma.";
RL FEBS LETT. 182:181-184(1985).
[4]
RN SEQUENCE OF 108-138.
RX MEDLINE; 84132557.
RA SEIDAH N.G., RAMASHARMA K., SAIRAM M.R., CHRETIEN M.;
RT "Partial amino acid sequence of a human seminal plasma peptide with
RL inhibin-like activity.";
RL FEBS LETT. 167:98-102(1984).
[5]
RN SEQUENCE OF 68-159.
RX MEDLINE; 85216629.
RA LI C.H., HAMMONDS R.G., RAMASHARMA K., CHUNG D.;
RT "Human seminal alpha inhibitors: isolation, characterization, and
RL structure.";
RL PROC. NATL. ACAD. SCI. U.S.A. 82:4041-4044(1985).
CC -1- FUNCTION: SEMENOGELIN I IS THE PREDOMINANT PROTEIN IN HUMAN SEMEN.
CC IT PARTICIPATES IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE
CC ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS
CC OF SEMENOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY
CC CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS
CC THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE
CC PROTEASE.
CC -1- FUNCTION: ALPHA-INHIBIN-92 AND ALPHA-INHIBIN-31, DERIVED FROM
CC THE PROTEOLYTIC DEGRADATION OF SEMENOGELIN, INHIBIT THE SECRETION
CC OF PITUITARY FOLLICLE-STIMULATING HORMONE.
CC -1- SUBUNIT: OCCURS IN DISULFIDE-LINKED COMPLEXES WHICH MAY ALSO
CC CONTAIN TWO LESS ABUNDANT 71- AND 76-KD SEMENOGELIN-RELATED
CC POLYPEPTIDES.
CC -1- TISSUE SPECIFICITY: SEMINAL VESICLE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04440; G338019; -;
DR EMBL; Z47556; E133812; -;
DR EMBL; M81650; G487420; -;
DR PIR; A03254; WTHUB.
DR PIR; A31489; A31489.
DR MIM; 182140; -;
KW SEMEN; SEMINAL VESICLE; REPEAT; SIGNAL; POLYMORPHISM.
FT SIGNAL 1 23
FT CHAIN 24 462
FT MOD_RES 24 24 SEMENOGELIN I.
FT PEPTIDE 24 24 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT PEPTIDE 68 159 ALPHA-INHIBIN-92.
FT PEPTIDE 108 138 ALPHA-INHIBIN-31.
FT PEPTIDE 108 159 SEMINAL BASIC PROTEIN.
FT DISULFID 239 239 INTERCHAIN.
FT REPEAT 174 215 42 AA REPEAT 1.
FT REPEAT 234 275 42 AA REPEAT 2.
FT REPEAT 282 339 58 AA REPEAT 1.
FT REPEAT 342 399 58 AA REPEAT 2.
FT REPEAT 414 455 42 AA REPEAT 3.
FT VARIANT 79 79 S -> T (LESS COMMON GENETIC VARIANT).
FT CONFLICT 423 423 K -> N (IN REF. 2).
SQ SEQUENCE 462 AA; 52131 MW; C87515C7 CRC32;
Query Match 85.7%; Score 42; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.50e+00;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 284 NKISYQ 289
QY 2 NKISYQ 7
RESULT 5
ID E314_ADE03 STANDARD; PRT; 136 AA.
AC P11315;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-AUG-1990 (REL. 15, LAST ANNOTATION UPDATE)
DE EARLY E3 15.3 KD PROTEIN.
OS HUMAN ADENOVIRUS TYPE 3.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE; MASTADENOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87219876.
RA SIGNAS C., AKUSJARVI G., PETERSSON U.;
RT "Region E3 of human adenoviruses; differences between the oncogenic
RL adenovirus-3 and the non-oncogenic adenovirus-2.";
RL GENE 50:173-184(1986).
[2]
RN SEQUENCE OF 128-136 FROM N.A.
RX MEDLINE; 85108162.
RA SIGNAS C., AKUSJARVI G., PETERSSON U.;
RT "Adenovirus 3 fiber polypeptide gene: implications for the structure
RL of the fiber protein.";
RL J. VIROL. 53:672-678(1985).
RN [3]
RP IDENTIFICATION OF PROTEIN.
RX MEDLINE; 90156523.
RA HORTON T.H., TOLLEFSON A.E., WOLD W.S.M., GOODING L.R.;
RT "A protein serologically and functionally related to the group C E3
RL 14,700-kilodalton protein is found in multiple adenovirus
RT serotypes.";
RL J. VIROL. 64:1250-1255(1990).
CC -1- FUNCTION: PROTECT VIRUS-INFECTED CELLS FROM TNF-INDUCED CYTOLYSIS.
CC -1- SIMILARITY: TO EQUIVALENT PROTEIN IN OTHER ADENOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15952; G209910; -;
DR EMBL; X01998; G583843; -;
DR EMBL; M12411; G209927; -;
DR PIR; A27623; ERAD29.
KW EARLY PROTEIN.
SQ SEQUENCE 136 AA; 15266 MW; 759E40EA CRC32;
Query Match 83.7%; Score 41; DB 1; Length 136;
Best Local Similarity 71.4%; Pred. No. 6.44e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 62 QSKISYE 68
QY 1 QNKISYQ 7
RESULT 6
ID BGLS_HANAN STANDARD; PRT; 825 AA.
AC P06835;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE BETA-GLUCOSIDASE PRECURSOR (EC 3.2.1.21) (GENTIOBIASE) (CELLOBIASE)
DE (BETA-D-GLUCOSIDE GLUCOHYDROLASE).

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 61 QAKISYE 67
1 QNKISYQ 7

RESULT 9
ID YUFO.ECOLI STANDARD; PRT; 142 AA.
AC P3297;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 13-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 16.0 KD PROTEIN IN AIDB-SGAT INTERGENIC REGION.
GN YUFO.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 95334362.
RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
RA BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL NUCLEIC ACIDS RES. 23:2105-2119(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14003; G537030; -;
DR EMBL; AE000491; G1790633; -;
DR ECGENE; EG12489; YUFO.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSEM 42 61 POTENTIAL.
SQ SEQUENCE 142 AA; 16034 MW; 78C83A3C CRC32;

Query Match 81.6%; Score 40; DB 1; Length 142;
Best Local Similarity 57.1%; Pred. No. 1.17e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 10 HSKIAYQ 16
1 QNKISYQ 7

RESULT 10
ID VH65.NPVAC STANDARD; PRT; 553 AA.
AC Q08539;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE EARLY 65 KD PROTEIN.
GN HB65.
OS AUTOGRAHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMPNV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E;
RX MEDLINE; 93381786.
RA BECKER D., KNEBEL-MOERSDORF D.;
RT "Sequence and temporal appearance of the early transcribed
RT baculovirus gene HB65.";
RL J. VIROL. 67:5867-5872(1993).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN-C6;
RX MEDLINE; 94303173.
RA AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERBER M., POSSEE R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL VIROLOGY 202:586-605(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X73577; G313680; -;
DR EMBL; L22858; G559174; -;
DR PIR; S35872; S35872.
KW EARLY PROTEIN.
SQ SEQUENCE 553 AA; 65576 MW; 420D4413 CRC32;

Query Match 81.6%; Score 40; DB 1; Length 553;
Best Local Similarity 83.3%; Pred. No. 1.17e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 315 QNKITY 320
1 QNKISY 6

RESULT 11
ID TOP1.MYCGE STANDARD; PRT; 709 AA.
AC P47388;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)
DE (UNWISTING ENZYME) (SWIVELASE).
GN TOPA OR MG122.
OS MYCOPLASMA GENITALIUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZMAN J.L., WEIDMAN J.F., SNALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL SCIENCE 270:397-403(1995).
RN [2]
RP SEQUENCE OF 399-481 AND 527-657 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 94075230.
RA PETERSON S.N., HU P.-C., BOTT K.F., HUTCHISON C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. BACTERIOL. 175:7918-7930(1993).
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA BACKBONE BOND, IT
CC SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN WHICH A TYROSYL OXYGEN
CC IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS AT ONE END OF THE
CC ENZYME-SEVERED DNA STRAND.
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39691; G1045802; -.
DR EMBL; U02134; G409912; -.
DR EMBL; U02242; G407259; -.
DR TIGR; MG122; -.
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
DR PFAM; PF01131; Topoisom_bac; 1.
DR HSP; P06612; IECL.
KW ISOMERASE; TOPOISOMERASE; DNA-BINDING.
FT ACT_SITE 334 334 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 709 AA; 82544 MW; 18063FDC CRC32;

Query Match 81.6%; Score 40; DB 1; Length 709;
Best Local Similarity 57.1%; Pred. No. 1.17e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 577 ENKVNQY 583
QY 1 QNKISYQ 7

RESULT 12
ID VM3_REOVD STANDARD; PRT; 719 AA.
AC P12419.
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE MAJOR NONSTRUCTURAL PROTEIN MU-NS.
GN M3.
OS REOVIRUS (TYPE 3 / STRAIN DEARING).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ORTHOREOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204900.
RA WIENER J.R., BARTLETT J.A., JOELIK W.K.;
RT "The sequences of reovirus serotype 3 genome segments M1 and M3
RT encoding the minor protein mu 2 and the major nonstructural protein
RT mu NS, respectively.";
RL VIROLOGY 169:293-304(1989).
CC -!- FUNCTION: THE MU-NONSTRUCTURAL PROTEIN IS A SS RNA-BINDING
CC PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27262; G333690; -.
DR PIR; B30179; MNXR3D.
KW NONSTRUCTURAL PROTEIN; RNA-BINDING.
FT MOD_RES 1 1 BLOCKED.
SQ SEQUENCE 719 AA; 79960 MW; 714DDF6E CRC32;

Query Match 81.6%; Score 40; DB 1; Length 719;
Best Local Similarity 71.4%; Pred. No. 1.17e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 658 QNKITYQ 664
QY 1 QNKISYQ 7
```

```
RESULT 13
ID FLGB_BACSU STANDARD; PRT; 129 AA.
AC P24500;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE FLAGELLAR BASAL-BODY ROD PROTEIN FLGB.
GN FLGB.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91285431.
RA ZUBERI A.R., YING C., BISCHOFF D.S., ORDAL G.W.;
RT "Gene-protein relationships in the flagellar hook-basal body complex
RT of Bacillus subtilis: sequences of the flgB, flgC, flgG, flgE and
RT flf genes.";
RL GENE 101:23-31(1991).
CC -!- SUBUNIT: THE BASAL BODY CONSTITUTES A MAJOR PORTION OF THE
CC FLAGELLAR ORGANELLE AND CONSISTS OF FOUR RINGS (L,P,S, AND M)
CC MOUNTED ON A CENTRAL ROD.
CC -!- SUBUNIT: THE ROD CONSISTS OF ABOUT 26 SUBUNITS OF FLGG IN THE
CC DISTAL PORTION, AND FLGB, FLGC AND FLGF ARE THOUGHT TO BUILD UP
CC THE PROXIMAL PORTION OF THE ROD WITH ABOUT 6 SUBUNITS EACH.
CC -!- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M54965; G142915; -.
DR EMBL; Z99112; E1185209; -.
DR PIR; JG0019; JG0019.
DR SUBTILIST; BG10237; FLGB.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
DR PFAM; PF00460; flg_bb_rod; 1.
KW FLAGELLA.
SQ SEQUENCE 129 AA; 14443 MW; 550868EA CRC32;

Query Match 79.6%; Score 39; DB 1; Length 129;
Best Local Similarity 71.4%; Pred. No. 2.10e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 102 QNKINYQ 108
QY 1 QNKISYQ 7

RESULT 14
ID XYL_R_BACSU STANDARD; PRT; 384 AA.
AC P16557;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE XYLLOSE REPRESSOR.
GN XYL_R.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=W23.
RX MEDLINE; 89291732.
RA KREUZER P., GAERTNER D., ALLMANSBERGER R., HILLEN W.;
RT "Identification and sequence analysis of the Bacillus subtilis W23
RT xylR gene and xyl operator.";
RL J. BACTERIOL. 171:3840-3845(1989).
```


CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF XYLOSE-UTILIZING ENZYMES.
CC -!- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M27248; G143841; -
CC FIR: A32885; A32885.
CC DR PROSITE: PS01125; ROK; 1.
CC DR PFAM: PF00480; ROK; 1.
CC KW TRANSCRIPTION REGULATION; DNA-BINDING; REPRESSOR; XYLOSE METABOLISM.
FT DNA_BIND 29 48 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 384 AA; 42295 MW; FEALAE85 CRC32;

Query Match 79.6%; Score 39; DB 1; Length 384;
Best Local Similarity 57.1%; Pred. No. 2.10e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 275 EKKVSYQ 281
:::|||||
QY 1 QNKISYQ 7

RESULT 15
ID ACH6_CAEEL STANDARD; PRT; 487 AA.
AC Q93149;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ACETYLCHOLINE RECEPTOR LIKE PROTEIN, BETA-TYPE SUBUNIT ACR-3
DE PRECURSOR.
DE GNR-3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNITEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BAYLIS H.A., MATSUDA K., SQUIRE M.D., FLEMING J.T., HARVEY R.J.,
RA DARLINGTON M.G., BARNARD E.A., SATTELLE D.B.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: POSSIBLE ACETYLCHOLINE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y08637; E274054; -
CC PROSITE: PS00236; NEUROTR_ION_CHANNEL; 1.
CC DR PFAM: PF00065; neur_chan; 1.
CC KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN;
KW TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 ?
FT CHAIN ? 487
FT DOMAIN ? 231
FT TRANSMEM 232 252
FT TRANSMEM 259 279
FT TRANSMEM 294 314
FT DOMAIN 315 439
FT TRANSMEM 440 460
FT CARBOHYD 46
POTENTIAL.
ACETYLCHOLINE RECEPTOR LIKE PROTEIN,
ALPHA-TYPE SUBUNIT ACR-3.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
POTENTIAL.
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
POTENTIAL.
POTENTIAL.

FT DISULFID 151 165 BY SIMILARITY.
SQ SEQUENCE 487 AA; 56245 MW; 102629B3 CRC32;

Query Match 79.6%; Score 39; DB 1; Length 487;
Best Local Similarity 71.4%; Pred. No. 2.10e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 214 RSKISYQ 220
:::|||||
QY 1 QNKISYQ 7

Search completed: Thu Oct 28 11:30:15 1999
Job time : 8 secs.

THIS PAGE BLANK (USPTO)

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:30:33 1999; MasPar time 4.13 Seconds
 Tabular output not generated. 92.427 Million cell updates/sec

Title: >US-09-081-707-11
 Description: (1-7) from US09081707.pap
 Perfect Score: 49

Sequence: 1 ONKISYQ 7

Scoring table: PAM 150
 Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrmb19

1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
 5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
 9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
 13:sp.vertebrate 14:sp.virus

Statistics: Mean 20.112; Variance 21.027; scale 0.956

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	100.0	287	2	085359 28 KDA MAJOR SURFACE A	2.94e-01
2	49	91.8	292	2	054429 PLASMID PIL2614 REPLIC	3.01e+00
3	45	91.8	704	6	029545 PORCINE INHIBITOR OF C	3.01e+00
4	43	87.8	51	2	053465 6 KDA ANTIGEN (FRAGMEN	9.17e+00
5	43	87.8	546	2	053850 GS60 ANTIGEN.	9.17e+00
6	43	87.8	664	6	002849 PEPTIDYLARGININE DEIMI	9.17e+00
7	42	85.7	433	5	061810 C18H7.8 PROTEIN.	1.58e+01
8	41	83.7	136	14	083121 ORF 15.3 KDA.	2.70e+01
9	41	83.7	175	14	065386 HYPOTHETICAL 20.1-KD P	2.70e+01
10	41	83.7	621	13	091192 MX1 PROTEIN.	2.70e+01
11	41	83.7	623	13	098991 MX2 PROTEIN.	2.70e+01
12	41	83.7	623	13	098992 MX3 PROTEIN.	2.70e+01
13	41	83.7	623	13	098990 MX1 PROTEIN.	2.70e+01
14	41	83.7	623	13	091197 RBTMX3.	2.70e+01
15	41	83.7	635	13	091196 RBTMX2.	2.70e+01
16	41	83.7	669	6	028085 CCP MODULES 3-12, WITH	2.70e+01
17	41	83.7	677	14	065799 VIRION SPIKE GLYCOPROT	2.70e+01
18	41	83.7	677	14	098953 VIRION SPIKE GLYCOPROT	2.70e+01
19	40	81.6	128	5	018126 T23G4.2 PROTEIN.	4.56e+01
20	40	81.6	135	14	067732 15.3 KDA PROTEIN.	4.56e+01

21	40	81.6	326	2	084563 YOP TRANSLOCATION J.	4.56e+01
22	40	81.6	344	2	099003 ADHESIN PRECURSOR.	4.56e+01
23	40	81.6	612	5	044048 ASPARAGINE-RICH PROTEI	4.56e+01
24	40	81.6	967	2	026050 TYPE III RESTRICTION E	4.56e+01
25	39	79.6	221	10	042376 GROUP 3 LEA PROTEIN MG	7.63e+01
26	39	79.6	222	8	034402 CYTOCHROME B (FRAGMENT	7.63e+01
27	39	79.6	222	8	036182 CYTOCHROME B (FRAGMENT	7.63e+01
28	39	79.6	258	5	001551 COSMID F40E3.	7.63e+01
29	39	79.6	282	10	065009 BURP DOMAIN CONTAINING	7.63e+01
30	39	79.6	332	5	082072 C25F9.8 PROTEIN.	7.63e+01
31	39	79.6	406	5	017460 HYPOTHETICAL 47.3 KD P	7.63e+01
32	39	79.6	425	6	018861 INTERPHOTORECEPTOR RET	7.63e+01
33	39	79.6	554	5	017459 HYPOTHETICAL 63.3 KD P	7.63e+01
34	39	79.6	605	5	010916 HYPOTHETICAL 70.1 KD P	7.63e+01
35	39	79.6	875	5	076698 TL7A3.1 PROTEIN.	7.63e+01
36	39	79.6	1021	2	025200 HYPOTHETICAL 120.1 KD	7.63e+01
37	39	79.6	1085	5	024363 I FACTOR.	7.63e+01
38	39	79.6	1190	11	062412 NEBULIN (FRAGMENT).	7.63e+01
39	39	79.6	1234	5	024690 HYPOTHETICAL 141.0 KD	7.63e+01
40	39	79.6	2178	2	046149 ALPHA-TOXIN.	7.63e+01
41	39	79.6	2567	13	090574 FILAMIN.	7.63e+01
42	39	79.6	2602	4	075369 BETA-FILAMIN.	7.63e+01
43	39	79.6	6669	4	Q15346 NEBULIN.	7.63e+01
44	38	77.6	680	2	053990 TBPB PRECURSOR.	1.26e+02
45	38	77.6	1299	4	060343 KIAA0603 PROTEIN.	1.26e+02

ALIGNMENTS

RESULT 1
 ID 085359 PRELIMINARY; PRT; 287 AA.
 AC 085359;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE 28 KDA MAJOR SURFACE ANTIGEN-1.
 OS EHRlichia CANIS.
 OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
 OC RICKETTSIAE; EHRlichiae; EHRlichia.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=OKLAHOMA;
 RX MEDLINE; 98321180.
 RA REDDY G.R., SULSONA C.R., BARBET A.F., MAHAN S.M., BURRIDGE M.J.,
 RA ALLEMAN A.R.;
 RT "Molecular characterization of a 28 kDa surface antigen gene family
 of the tribe Ehrlichiae."
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 247:636-643(1998).
 DR EMBL; AF062762; G3327965; .
 SQ SEQUENCE 287 AA; 32014 MW; 1216502E CRC32;

Query Match 100.0%; Score 49; DB 2; Length 287;
 Best Local Similarity 100.0%; Pred. No. 2.94e-01;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 214 ONKISYQ 220

Qy 1 ONKISYQ 7

RESULT 2
 ID 054429 PRELIMINARY; PRT; 292 AA.
 AC 054429;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PLASMID PIL2614 REPLICATION PROTEIN
 DE (REP), TYPE IC RESTRICTION SUBUNIT
 DE (HSDR), TYPE IC MODIFICATION SUBUNIT
 DE (HSDM), TYPE IC MODIFICATION SUBUNIT
 DE (HSDS), AND PHAGE ABORTIVE INFECTION PROTEIN
 DE (ABI420) GENES, COMPLETE CDS (HSDR) (HSDS) (ABI420).
 OS LACTOCOCCUS LACTIS (STREPTOCOCCUS LACTIS).

Mon Nov 1 12:18:28 1999

```

OG PLASMID PIL2614.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC LACTOCOCCUS.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE; 98101482.
RA SCHOUER C., CLIER F., LERAYER A.L., EHRLICH S.D., CHOPIN M.C.;
RT "A type IC restriction-modification system in Lactococcus lactis.";
RL J. BACTERIOL. 180:407-411(1998).
DR EMBL; U90222; G2865241; -.
KW PLASMID.
SQ SEQUENCE 292 AA; 34223 MW; 9EBE797D CRC32;

Query Match 91.8%; Score 45; DB 2; Length 292;
Best Local Similarity 85.7%; Pred. No. 3.01e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 23 QNKISYE 29
|:|:|:|:|:
QY 1 QNKISYQ 7

RESULT 3
ID Q29545 PRELIMINARY; PRT; 704 AA.
AC Q29545;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PORCINE INHIBITOR OF CARBONIC ANHYDRASE PRECURSOR.
GN PICA.
OS SEQUENCE FROM N.A.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 93099129.
RA ROUSH E.D., FIERKE C.A.;
RT "Purification and characterization of a carbonic anhydrase II
inhibitor from porcine plasma.";
RL BIOCHEMISTRY 31:12536-12542(1992).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 97254619.
RA WUEBBENS M.W., ROUSH E.D., DECASTRO C.M., FIERKE C.A.;
RT "Cloning, sequencing, and recombinant expression of the porcine
inhibitor of carbonic anhydrase: a novel member of the transferrin
family.";
RL BIOCHEMISTRY 36:4327-4336(1997).
CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
DR EMBL; U36916; G1016330; -.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 1.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
DR PFAM; PF00405; transferrin; 2.
DR SIGNAL; IRON TRANSPORT; GLYCOPROTEIN; METAL-BINDING.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 704 PORCINE INHIBITOR OF CARBONIC ANHYDRASE.
SQ SEQUENCE 704 AA; 77634 MW; 2E961A99 CRC32;

Query Match 91.8%; Score 45; DB 6; Length 704;
Best Local Similarity 71.4%; Pred. No. 3.01e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 667 QDKITYQ 673
|:|:|:|:|
QY 1 QNKISYQ 7

RESULT 4

```

```

ID Q53465 PRELIMINARY; PRT; 51 AA.
AC Q53465;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 6 KDA ANTIGEN (FRAGMENT).
OS PASTEURILLA HAEMOLYTICA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC PASTURELLA.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 95027654.
RA WELDON S.K., MOSIER D.A., SIMONS K.R., CRAVEN R.C., CONFER A.W.;
RT "Identification of a potentially important antigen of Pasteurella
haemolytica.";
RL VET. MICROBIOL. 40:283-291(1994).
DR EMBL; S74144; E155751; -.
FT NON_TER 1 1
FT NON_TER 51 51
SQ SEQUENCE 51 AA; 5954 MW; 93E7B40E CRC32;

Query Match 87.8%; Score 43; DB 2; Length 51;
Best Local Similarity 71.4%; Pred. No. 9.17e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 12 EDKISYQ 18
|:|:|:|:|
QY 1 QNKISYQ 7

RESULT 5
ID Q51850 PRELIMINARY; PRT; 546 AA.
AC Q51850;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE GS60 ANTIGEN.
GN GS60.
OS PASTURELLA HAEMOLYTICA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC PASTEURILLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE 1;
RA LO R.Y.C., MELLORES A.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U42028; G1353671; -.
SQ SEQUENCE 546 AA; 60815 MW; 84B6A9DA CRC32;

Query Match 87.8%; Score 43; DB 2; Length 546;
Best Local Similarity 71.4%; Pred. No. 9.17e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 57 EDKISYQ 53
|:|:|:|:|
QY 1 QNKISYQ 7

RESULT 6
ID Q02849 PRELIMINARY; PRT; 664 AA.
AC Q02849;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PEPTIDYLARGININE DEIMINASE (EC 3.5.3.15)
DE (PROTEIN-ARGININE DEIMINASE).
GN PAD.
OS OVIS ARIES (SHEEP).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; CAPRINAE; OVIS.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-WOOL FOLLICLES.
RC

```

```
RA ROGERS G.E., WINTER B.P., POWELL B.C., NESCI A.N.;
RL J. INVEST. DERMATOL. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: PROTEIN L-ARGININE + H(2)O = PROTEIN
CC L-CITRULLINE + NH(3).
DR EMBL; U85264; G2055390; -.
KW HYDROLASE.
SQ SEQUENCE 664 AA; 74747 MW; 6813DB1B CRC32;

Query Match      87.8%; Score 43; DB 6; Length 664;
Best Local Similarity 71.4%; Pred. No. 9.17e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 232 QNKVSYE 238
   |||:|
   1 QNKISYQ 7

RESULT 7
ID O61810; PRELIMINARY; PRT; 433 AA.
AC O61810;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE C18H7.8 PROTEIN.
GN C18H7.8
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA TIR-WOLLAM A., FRONICK W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF067607; G3165526; -.
SQ SEQUENCE 433 AA; 50343 MW; 47266C20 CRC32;

Query Match      85.7%; Score 42; DB 5; Length 433;
Best Local Similarity 71.4%; Pred. No. 1.58e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 180 QNKITFQ 186
   |||:|
   1 QNKISYQ 7

RESULT 8
ID Q83121; PRELIMINARY; PRT; 136 AA.
AC Q83121;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE ORF 15.3 KDA.
OS MASTADENOVIRUS.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=87-922;
RX MEDLINE; 96146733.
RA KAJON A.E., WADELL G.;
RT "Sequence analysis of the E3 region and fiber gene of human
RT adenovirus genome type 7h."
RL VIROLOGY 215:190-196(1996).
DR EMBL; Z48954; G762864; -.
SQ SEQUENCE 136 AA; 15275 MW; 87E42674 CRC32;

Query Match      83.7%; Score 41; DB 14; Length 136;
Best Local Similarity 71.4%; Pred. No. 2.70e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 62 QSKISYE 68
   |||:|
   1 QNKISYQ 7

RESULT 9
ID O65386; PRELIMINARY; PRT; 175 AA.
AC O65386;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 20.1 KD PROTEIN.
OC BANANA BUNCHY TOP VIRUS.
OS VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95302043.
RA BURNS T.M., HARDING R.M., DALE J.L.;
RT "The genome organization of banana bunchy top virus: analysis of six
RT ssDNA components."
RL J. GEN. VIROL. 76:1471-1482(1995).
DR EMBL; L41574; G867679; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 175 AA; 20111 MW; 48C78A21 CRC32;

Query Match      83.7%; Score 41; DB 14; Length 175;
Best Local Similarity 57.1%; Pred. No. 2.70e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 168 QNRITYH 174
   |||:|
   1 QNKISYQ 7

RESULT 10
ID Q91192; PRELIMINARY; PRT; 621 AA.
AC Q91192;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MX1 PROTEIN.
OS ONCORHYNCHUS MYKISS (RAINBOW TROUT) (SALMO GAIARDNERI).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; SALMONIFORMES; SALMONIDAE;
OC ONCORHYNCHUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96020503.
RA TROBRIDGE G.D., LEONG J.A.;
RT "Characterization of a rainbow trout Mx gene."
RL J. INTERFERON CYTOKINE RES. 15:691-702(1995).
DR EMBL; U30253; G905389; -.
DR PROSITE; PS00410; DYNAMIN; 1.
DR PFAM; PF00350; dynamin; 1.
DR PFAM; PF01031; dynamin_2; 1.
```

```
SQ SEQUENCE 621 AA; 70640 MW; 8F350554 CRC32;
Query Match 83.7%; Score 41; DB 13; Length 621;
Best Local Similarity 71.4%; Pred. No. 2.70e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 85 HGKISYQ 91
: |||||
QY 1 QNKISYQ 7

RESULT 11
ID Q98991 PRELIMINARY; PRT; 623 AA.
AC Q98991;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MX2 PROTEIN.
OS SALMO SALAR (ATLANTIC SALMON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; SALMONIFORMES; SALMONIDAE;
OC SALMO.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBERTSEN B., TROBRIDGE G.D., LEONG J.C.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U66476; G1519386; -.
DR PROSITE; PS00410; DYNAMIN; 1.
DR PFAM; PF00350; dynamin; 1.
DR PFAM; PF01031; dynamin_2; 1.
DR PFAM; PF01031; dynamin_2; 1.
SQ SEQUENCE 623 AA; 70804 MW; B18012DA CRC32;

Query Match 83.7%; Score 41; DB 13; Length 623;
Best Local Similarity 71.4%; Pred. No. 2.70e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 85 HGKISYQ 91
: |||||
QY 1 QNKISYQ 7

RESULT 12
ID Q98992 PRELIMINARY; PRT; 623 AA.
AC Q98992; P79904;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE MX3 PROTEIN.
OS SALMO SALAR (ATLANTIC SALMON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; SALMONIFORMES; SALMONIDAE;
OC SALMO.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBERTSEN B., TROBRIDGE G.D., LEONG J.C.;
RL "Cloning of the Mx Genes of Atlantic Salmon (Salmo salar).";
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DR EMBL; U66477; G1783385; -.
DR PROSITE; PS00410; DYNAMIN; 1.
DR PFAM; PF00350; dynamin; 1.
DR PFAM; PF01031; dynamin_2; 1.
DR PFAM; PF01031; dynamin_2; 1.
SQ SEQUENCE 623 AA; 70923 MW; DAC2A40A CRC32;

Query Match 83.7%; Score 41; DB 13; Length 623;
Best Local Similarity 71.4%; Pred. No. 2.70e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 85 HGKISYQ 91
: |||||
QY 1 QNKISYQ 7

RESULT 13
ID Q98990 PRELIMINARY; PRT; 623 AA.
AC Q98990;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MX1 PROTEIN.
OS SALMO SALAR (ATLANTIC SALMON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; SALMONIFORMES; SALMONIDAE;
OC SALMO.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBERTSEN B., TROBRIDGE G.D., LEONG J.C.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U66475; G1519384; -.
DR PROSITE; PS00410; DYNAMIN; 1.
DR PFAM; PF00350; dynamin; 1.
DR PFAM; PF01031; dynamin_2; 1.
DR PFAM; PF01031; dynamin_2; 1.
SQ SEQUENCE 623 AA; 70895 MW; 99E0B5A5 CRC32;

Query Match 83.7%; Score 41; DB 13; Length 623;
Best Local Similarity 71.4%; Pred. No. 2.70e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 85 HGKISYQ 91
: |||||
QY 1 QNKISYQ 7

RESULT 14
ID Q91197 PRELIMINARY; PRT; 623 AA.
AC Q91197;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RBTMX3.
OS ONCORHYNCHUS MYKISS (RAINBOW TROUT) (SALMO GAIRDNERI).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; SALMONIFORMES; SALMONIDAE;
OC ONCORHYNCHUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97323265.
RA TROBRIDGE G.D., CHIOU P.P., LEONG J.C.;
RT "Cloning of the rainbow trout (Oncorhynchus mykiss) Mx2 and Mx3 cDNAs
and characterization of trout Mx protein expression in salmon
cells.";
RL J. VIROL. 71:5304-5311(1997).
DR EMBL; U47946; G1399454; -.
DR PROSITE; PS00410; DYNAMIN; 1.
DR PFAM; PF00350; dynamin; 1.
DR PFAM; PF01031; dynamin_2; 1.
DR PFAM; PF01031; dynamin_2; 1.
SQ SEQUENCE 623 AA; 70962 MW; 8320E073 CRC32;

Query Match 83.7%; Score 41; DB 13; Length 623;
Best Local Similarity 71.4%; Pred. No. 2.70e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 85 HGKISYQ 91
: |||||
QY 1 QNKISYQ 7

RESULT 15
ID Q91196 PRELIMINARY; PRT; 635 AA.
AC Q91196;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RBTMX2.
OS ONCORHYNCHUS MYKISS (RAINBOW TROUT) (SALMO GAIRDNERI).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; SALMONIFORMES; SALMONIDAE;
```

OC ONCORHYNCHUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97332365.
RA TROBRIDGE G.D., CHIOU P.P., LEONG J.C.;
RT "Cloning of the rainbow trout (Oncorhynchus mykiss) Mx2 and Mx3 cDNAs
RT and characterization of trout Mx protein expression in salmon
RT cells.";
RL J. VIROL. 71:5304-5311(1997).
DR EMBL: U47945; GI399452; -;
DR PROSITE: PS00410; DYNAMIN; 1.
DR PFAM: PF00350; dynamin; 1.
DR PFAM: PF01031; dynamin_2; 1.
SQ SEQUENCE 635 AA; 72112 MW; 4015A5A9 CRC32;

Query Match 83.7%; Score 41; DB 13; Length 635;
Best Local Similarity 71.4%; Pred. No. 2.70e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 85 HGKISYQ 91
QY 1 QNKISYQ 7
:|||||

Search completed: Thu Oct 28 11:31:17 1999
Job time : 44 secs.

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run On: Thu Oct 28 11:33:40 1999; MasPar time 3.26 Seconds
Tabular output not generated. 45.628 Million cell updates/sec

Title: >US-09-081-707-12
(1-7) from US09081707 pep
Perfect Score: 48
Sequence: 1 ENKISYQ 7

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 14.768; Variance 35.751; scale 0.413

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	47	97.9	17	17	Prostate specific ant	2.19e+01
2	47	97.9	17	26	Oligopeptide 19 cleav	2.19e+01
3	45	93.8	28	32	Ehrlichia canis VSA1	4.06e+01
4	42	87.5	7	17	Prostate specific ant	1.01e+02
5	42	87.5	7	26	Oligopeptide 29 cleav	1.01e+02
6	42	87.5	8	26	Oligopeptide 126 base	1.01e+02
7	42	87.5	8	26	Oligopeptide 132 base	1.01e+02
8	42	87.5	8	17	Prostate specific ant	1.01e+02
9	42	87.5	8	17	Prostate specific ant	1.01e+02
10	42	87.5	9	17	Prostate specific ant	1.01e+02
11	42	87.5	9	17	Prostate specific ant	1.01e+02
12	42	87.5	9	26	Oligopeptide 132 base	1.01e+02
13	42	87.5	9	26	Oligopeptide 68 cleav	1.01e+02
14	42	87.5	10	26	Oligopeptide 28 cleav	1.01e+02
15	42	87.5	10	26	Oligopeptide 9 cleave	1.01e+02
16	42	87.5	10	26	Oligopeptide 43 cleav	1.01e+02

17	42	87.5	10	26	W33437	Oligopeptide 17 cleav	1.01e+02
18	42	87.5	10	26	W33500	Oligopeptide 80 cleav	1.01e+02
19	42	87.5	10	17	R89626	Prostate specific ant	1.01e+02
20	42	87.5	10	17	R89600	Prostate specific ant	1.01e+02
21	42	87.5	10	17	R89663	Prostate specific ant	1.01e+02
22	42	87.5	10	17	R89592	Prostate specific ant	1.01e+02
23	42	87.5	11	17	R89657	Prostate specific ant	1.01e+02
24	42	87.5	11	17	R89657	Prostate specific ant	1.01e+02
25	42	87.5	11	17	R89657	Prostate specific ant	1.01e+02
26	42	87.5	11	17	R89657	Prostate specific ant	1.01e+02
27	42	87.5	11	17	R89657	Prostate specific ant	1.01e+02
28	42	87.5	12	26	W33493	Oligopeptide 73 cleav	1.01e+02
29	42	87.5	12	26	W33497	Oligopeptide 77 cleav	1.01e+02
30	42	87.5	12	26	W33499	Oligopeptide 80 based	1.01e+02
31	42	87.5	12	26	W33496	Oligopeptide 76 based	1.01e+02
32	42	87.5	12	17	R89662	Prostate specific ant	1.01e+02
33	42	87.5	12	17	R89660	Prostate specific ant	1.01e+02
34	42	87.5	12	26	W33492	Oligopeptide 72 cleav	1.01e+02
35	42	87.5	12	17	R89658	Prostate specific ant	1.01e+02
36	42	87.5	12	17	R89656	Prostate specific ant	1.01e+02
37	42	87.5	12	17	R89655	Prostate specific ant	1.01e+02
38	42	87.5	12	17	R89652	Prostate specific ant	1.01e+02
39	42	87.5	12	26	W33489	Oligopeptide 69 cleav	1.01e+02
40	42	87.5	12	17	R89593	Prostate specific ant	1.01e+02
41	42	87.5	14	26	W33495	Oligopeptide 75 based	1.01e+02
42	42	87.5	25	26	W33425	Oligopeptide 5 cleave	1.01e+02
43	42	87.5	25	17	R89649	Prostate specific ant	1.01e+02
44	42	87.5	462	17	R89583	Human semenogelin I.	1.01e+02
45	42	87.5	462	26	W33420	Human semenogelin I.	1.01e+02

ALIGNMENTS

RESULT 1
ID R89602 standard; peptide: 17 AA.
AC R89602;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
KW gel structure; dissolution; prostate specific antigen; proteolysis;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
PN W09600503-Al.
PD 11-JAN-1996.
PF 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI) MERCK & CO INC.
PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR WPI; 96-077275/08.
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Disclosure; Page 60; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSI and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC action of prostate specific antigen (PSA), a protease with
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond
CC or peptide linker, to a cytotoxic agent and used to treat
CC prostate cancer.
SQ Sequence 17 AA;

Query Match 97.9%; Score 47; DB 17; Length 17;
Best Local Similarity 85.7%; Pred. No. 2.19e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Db      3 dnkisyq 9
        :|||||
QY      1 ENKISYQ 7

RESULT  2
ID      W33439 standard; Peptide; 17 AA.
AC      W33439;
DT      27-MAR-1998 (first entry)
DE      Oligopeptide 19 cleaved by free prostate specific antigen.
KW      Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
KW      cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
KW      adenocarcinoma; prostate metastases; prostate cancer; treatment;
KW      detection; cytotoxic conjugate; activation; quantitative assay.
OS      Synthetic.
FH      Key Location/Qualifiers
FT      Cleavage_site 9..10 /note= "PSA specific cleavage site"
FT      PN WO9712624-A1.
PD      10-APR-1997.
PF      02-OCT-1996; U15713.
PR      06-OCT-1995; US-540412.
PA      (MERI ) MERCK & CO INC.
PI      DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR      WPI; 97-225974/20...
DT      Oligopeptide(s) recognised and cleaved by free prostate specific
PT      antigen - useful in assays for active antigen, and in
PT      Oligopeptide-drug conjugates for prostatic cancer treatment
PS      Disclosure; Page 81; 192pp; English.
CC      The present sequence is a novel oligopeptide that is recognised and
CC      proteolytically cleaved by free prostate specific antigen (PSA). It was
CC      designed based on the sequences surrounding the PSA cleavage sites of
CC      Semenogelin I, a major sperm entrapping gel protein. PSA, which has
CC      chymotrypsin-like specificity is responsible for dissolution of the gel
CC      structure formed at ejaculation by proteolysis of the major gel proteins
CC      (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
CC      releases progressively motile spermatozoa. PSA complexed to
CC      alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
CC      measurements of PSA are useful for monitoring treatment of adenocarcinoma
CC      of the prostate. Prostate metastases are also known to secrete
CC      immunologically reactive PSA since serum PSA is detectable at high levels
CC      in prostatectomised patients showing widespread metastatic prostate
CC      cancer. Cytotoxic compounds that could be activated by the proteolytic
CC      activity of PSA should also be prostate cell specific as well as specific
CC      for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
CC      and this oligopeptide are useful in treatment of prostate cancer. The
CC      oligopeptide can also be used in a quantitative assay for enzymatically
CC      active PSA.
SQ      Sequence 17 AA;

Query Match 97.9%; Score 47; DB 26; Length 17;
Best Local Similarity 85.7%; Pred. No. 2.19e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db      3 dnkisyq 9
        :|||||
QY      1 ENKISYQ 7

RESULT  3
ID      W31096 standard; Protein; 287 AA.
AC      W31096;
DT      14-SEP-1998 (first entry)
DE      Ehrlichia canis VSA1 protein (partial sequence).
KW      MAP1 homologue; variable surface antigen; VSA1; rickettsia;
KW      DNA vaccine.
OS      Ehrlichia canis.
FH      Key Location/Qualifiers
FT      Peptide 1..25 /note= "putative signal peptide"
FT      PN WO9816554-A1.
PD      22-APR-1998.

```

```

PF      17-OCT-1997; U19044.
PR      17-OCT-1996; US-733230.
PA      (UYFL ) UNIV FLORIDA.
PI      Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC,
PI      Nyika A, Rurangirwa FR;
DR      WPI; 98-251232/22.
DR      N-PSDB; V07180.
DT      Composition containing nucleic acid encoding rickettsial antigen -
PT      useful for, e.g. stimulating protective immune response in humans or
PT      animals
PS      Claim 3; Fig 2C; 39pp; English.
CC      This is the full-length variable surface antigen VSA1 protein of
CC      Ehrlichia canis. Its amino acid sequence was deduced from open
CC      reading frame (ORF) 1 of a genomic locus (see V07180) of E. canis
CC      that was obtained on the basis of homology to the major antigenic
CC      protein MAP1 (see W51088) of Cowdria ruminantium. This genomic
CC      locus included 2 ORFs encoding similar, but non-identical proteins
CC      (see W51096-97). A claimed composition comprises a nucleic acid
CC      (see V07176-82) encoding a polypeptide (see W51088-99) that elicits
CC      a protective immune response against a rickettsial pathogen. The
CC      nucleic acid is used, in human or veterinary medicine, in vaccines
CC      to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
CC      species. The Ehrlichia antigenic polypeptides can be used
CC      diagnostically to detect antibodies associated with Ehrlichia
CC      infection (claimed).
SQ      Sequence 287 AA;

Query Match 93.8%; Score 45; DB 32; Length 287;
Best Local Similarity 85.7%; Pred. No. 4.06e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db      214 qnkisyq 220
        :|||||
QY      1 ENKISYQ 7

RESULT  4
ID      R89612 standard; peptide; 7 AA.
AC      R89612;
DT      02-SEP-1996 (first entry)
DE      Prostate specific antigen, semenogelin derived, cleavage substrate.
KW      Human; semenogelin I; sperm entrapping gel; ejaculation; prostate;
KW      gel structure; dissolution; prostate specific antigen; proteolysis;
KW      chymotrypsin like specificity; peptide substrate; cleavage site;
KW      assay; determination; proteolytic activity; identification;
KW      inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS      Homo sapiens.
FH      Key Location/Qualifiers
FT      Cleavage_site 6..7 /note= "prostate specific antigen proteolytic
FT      cleavage site"
FT      PN WO9600503-A1.
PD      11-JAN-1996.
PF      07-JUN-1995; U08156.
PR      28-JUN-1994; US-267092.
PR      15-MAR-1995; US-404833.
PA      (MERI ) MERCK & CO INC.
PI      DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR      WPI; 96-077275/08.
DT      New peptide substrates cleaved by prostate-specific antigen - also
PT      cytotoxic conjugates for treating prostate cancer, and assay for
PT      determination of PSA activity
PS      Claim 6; Page 64; 142pp; English.
CC      Human semenogelin I (hSI) is one of the major proteins, including
CC      hSII and fibronectin, in the sperm entrapping gel formed at
CC      ejaculation. This gel structure undergoes dissolution via the
CC      action of prostate specific antigen (PSA), a protease with
CC      chymotrypsin like specificity, which proteolyses the above major
CC      proteins. New substrates, including the present peptide, cleaved
CC      by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC      in assays to determine the proteolytic activity of free PSA in a
CC      sample, and to identify cpds. which inhibit the proteolytic
CC      activity of PSA, they may also be conjugated, via a covalent bond

```

CC or peptide linker, to a cytotoxic agent and used to treat
 CC prostate cancer.
 SQ Sequence 7 AA;

Query Match 87.5%; Score 42; DB 17; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 5

ID W33449 standard; Peptide; 7 AA.
 AC W33449;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 29 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 6..7 /note= "PSA specific cleavage site"
 PN WO9712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; UI5713.
 PR 06-OCT-1995; US-540412.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI: 97-225974/20.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 7 AA;

Query Match 87.5%; Score 42; DB 26; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 6

ID W33546 standard; Peptide; 8 AA.
 AC W33546;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 126 based on Semenogelin I cleavage site.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;

KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 6..7 /note= "PSA specific cleavage site"
 PN WO9712624-A1.
 PD 10-APR-1997.
 PF 02-OCT-1996; UI5713.
 PR 06-OCT-1995; US-540412.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 DR WPI: 97-225974/20.
 CC The present sequence is a novel oligopeptide designed based on the
 CC sequences surrounding the prostate specific antigen (PSA) cleavage sites
 CC of Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 8 AA;

Query Match 87.5%; Score 42; DB 26; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 7

ID W33435 standard; Peptide; 8 AA.
 AC W33435;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 15 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Modified_site 1 /note= "optionally N-terminally acetylated, when
 FT C-terminally amidated"
 FT Cleavage_site 6..7 /note= "PSA specific cleavage site"
 FT Modified_site 8 /note= "this Ser is optionally attached via a peptide
 FT bond to the 3'-amino group on the sugar ring of
 FT doxorubicin, alternatively it is optionally
 FT amidated when the peptide is N-terminally
 FT acetylated"

WO9712624-A1.
 10-APR-1997.
 02-OCT-1996; UI5713.
 06-OCT-1995; US-540412.

PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 PR WPI: 97-225974/20.
 PT Oligopeptide(s) recognised and cleaved by free prostate specific
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 3 and 7 and 17: Page 160: 142pp; English.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA. In a study of cleavage affinity, the IPA salt of this peptide
 CC (with an N-terminal acetyl group and a C-terminal amide group) was
 CC digested with enzymatically active PSA. After four hours 30 per cent of
 CC the peptide salt was cleaved.
 SQ Sequence 8 AA;

Query Match 87.5%; Score 42; DB 26; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 8
 ID R89598 standard; peptide; 8 AA.
 AC R89598;
 DT 02-SEP-1996 (first entry)
 DE Prostate specific antigen, semenogelin derived, cleavage substrate.
 KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
 KW gel structure; dissolution; prostate specific antigen; proteolysis;
 KW chymotrypsin like specificity; peptide substrate; cleavage site;
 KW assay; determination; proteolytic activity; identification;
 KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
 OS Homo.sapiens.
 FH Key Location/Qualifiers
 FT cleavage_site 6..7
 FT /note= "prostate specific antigen proteolytic
 FT cleavage site"
 FT modified_site 1
 FT /note= "opt. acylated"
 FT modified_site 12
 FT /note= "opt. amidated"
 FT WO9600503-A1.
 PN 11-JAN-1996.
 PF 07-JUN-1995; U08156.
 PR 28-JUN-1994; US-267092.
 PR 15-MAR-1995; US-404833.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 PR WPI: 96-077275/08.
 PT New peptide substrates cleaved by prostate-specific antigen - also
 PT cytotoxic conjugates for treating prostate cancer, and assay for
 PT determination of PSA activity
 PS Claim 3; Page 58; 142pp; English.
 CC Human semenogelin I (hSI) is one of the major proteins, including
 CC hSII and fibronectin, in the sperm entrapping gel formed at

CC ejaculation. This gel structure undergoes dissolution via the
 CC action of prostate specific antigen (PSA), a protease with
 CC chymotrypsin like specificity, which proteolyses the above major
 CC proteins. New substrates, including the present peptide, cleaved
 CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
 CC in assays to determine the proteolytic activity of free PSA in a
 CC sample, and to identify cpds. which inhibit the proteolytic
 CC activity of PSA, they may also be conjugated, via a covalent bond
 CC or peptide linker, to a cytotoxic agent and used to treat
 CC prostate cancer. In a PSA hydrolysis assay, the percentage of
 CC the present peptide cleaved by YORK PSA after 4 hrs. was 30 %
 CC acylated and amidated.
 SQ Sequence 8 AA;

Query Match 87.5%; Score 42; DB 17; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 9
 ID R89712 standard; peptide; 8 AA.
 AC R89712;
 DT 02-SEP-1996 (first entry)
 DE Prostate specific antigen, semenogelin derived, cleavage substrate.
 KW Human; semenogelin I; sperm entrapping gel; ejaculation; protease;
 KW gel structure; dissolution; prostate specific antigen; proteolysis;
 KW chymotrypsin like specificity; peptide substrate; cleavage site;
 KW assay; determination; proteolytic activity; identification;
 KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cleavage_site 6..7
 FT /note= "prostate specific antigen proteolytic
 FT cleavage site"
 FT WO9600503-A1.
 PN 11-JAN-1996.
 PF 07-JUN-1995; U08156.
 PR 28-JUN-1994; US-267092.
 PR 15-MAR-1995; US-404833.
 PA (MERI) MERCK & CO INC.
 PI DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 PR WPI: 96-077275/08.
 PT New peptide substrates cleaved by prostate-specific antigen - also
 PT cytotoxic conjugates for treating prostate cancer, and assay for
 PT determination of PSA activity
 PS Claim 3; Page 108; 142pp; English.
 CC Human semenogelin I (hSI) is one of the major proteins, including
 CC hSII and fibronectin, in the sperm entrapping gel formed at
 CC ejaculation. This gel structure undergoes dissolution via the
 CC action of prostate specific antigen (PSA), a protease with
 CC chymotrypsin like specificity, which proteolyses the above major
 CC proteins. New substrates, including the present peptide, cleaved
 CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
 CC in assays to determine the proteolytic activity of free PSA in a
 CC sample, and to identify cpds. which inhibit the proteolytic
 CC activity of PSA, they may also be conjugated, via a covalent bond
 CC or peptide linker, to a cytotoxic agent and used to treat
 CC prostate cancer.
 SQ Sequence 8 AA;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

Query Match 87.5%; Score 42; DB 17; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

```

RESULT 10
ID R89599 standard; peptide; 9 AA.
AC R89599;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; prostate;
KW gel structure; dissolution; prostate specific antigen; proteolysis;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 6.7
FT /note= "prostate specific antigen proteolytic
FT cleavage site"
PD WO9600503-A1.
PN 11-JAN-1996.
PR 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI ) MERCK & CO INC.
PI Defeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR WPI: 96-077275/08.
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Claim 3; Page 58; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSII and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond
CC or peptide linker, to a cytotoxic agent and used to treat
CC prostate cancer. In a PSA hydrolysis assay, the percentage of
CC the present peptide cleaved by YORK PSA after 4 hrs. was 55 %.
CC Sequence 9 AA;
Query Match 87.5%; Score 42; DB 17; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.01e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
QY 2 NKISYQ 7

RESULT 11
ID R89718 standard; peptide; 9 AA.
AC R89718;
DT 02-SEP-1996 (first entry)
DE Prostate specific antigen, semenogelin derived, cleavage substrate.
KW Human; semenogelin I; sperm entrapping gel; ejaculation; prostate;
KW gel structure; dissolution; prostate specific antigen; proteolysis;
KW chymotrypsin like specificity; peptide substrate; cleavage site;
KW assay; determination; proteolytic activity; identification;
KW inhibitor; cytotoxic agent; conjugated; treatment; prostate cancer.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cleavage_site 7.8
FT /note= "prostate specific antigen proteolytic
FT cleavage site"
PD WO9600503-A1.
PN 11-JAN-1996.
PR 07-JUN-1995; U08156.
PR 28-JUN-1994; US-267092.
PR 15-MAR-1995; US-404833.
PA (MERI ) MERCK & CO INC.
PI Defeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;

```

```

DR WPI: 96-077275/08.
PT New peptide substrates cleaved by prostate-specific antigen - also
PT cytotoxic conjugates for treating prostate cancer, and assay for
PT determination of PSA activity
PS Claim 4; Page 111; 142pp; English.
CC Human semenogelin I (hSI) is one of the major proteins, including
CC hSII and fibronectin, in the sperm entrapping gel formed at
CC ejaculation. This gel structure undergoes dissolution via the
CC action of prostate specific antigen (PSA), a protease with
CC chymotrypsin like specificity, which proteolyses the above major
CC proteins. New substrates, including the present peptide, cleaved
CC by PSA, i.e. peptides contg. a hSI PSA cleavage site, can be used
CC in assays to determine the proteolytic activity of free PSA in a
CC sample, and to identify cpds. which inhibit the proteolytic
CC activity of PSA, they may also be conjugated, via a covalent bond
CC or peptide linker, to a cytotoxic agent and used to treat
CC prostate cancer.
CC Sequence 9 AA;
Query Match 87.5%; Score 42; DB 17; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.01e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 nkisyq 7
QY 2 NKISYQ 7

RESULT 12
ID W33552 standard; Peptide; 9 AA.
AC W33552;
DT 27-MAR-1998 (first entry)
DE Oligopeptide 132 based on Semenogelin I cleavage site.
KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
KW detection; cytotoxic conjugate; activation; quantitative assay.
OS Synthetic.
FH Key Location/Qualifiers
FT cleavage_site 7.8
FT /note= "PSA specific cleavage site"
PD WO9712624-A1.
PN 10-APR-1997.
PR 02-OCT-1996; U15713.
PR 06-OCT-1995; US-540412.
PA (MERI ) MERCK & CO INC.
PI Defeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
DR WPI: 97-225974/20.
PT Oligopeptide(s) recognised and cleaved by free prostate specific
PT antigen - useful in assays for active antigen, and in
PT oligopeptide-drug conjugates for prostatic cancer treatment
PS Claim 4; Page 161; 192pp; English.
CC The present sequence is a novel oligopeptide designed based on the
CC sequences surrounding the prostate specific antigen (PSA) cleavage sites
CC of Semenogelin I, a major sperm entrapping gel protein. PSA, which has
CC chymotrypsin-like specificity is responsible for dissolution of the gel
CC structure formed at ejaculation by proteolysis of the major gel proteins
CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
CC releases progressively motile spermatozoa. PSA complexed to
CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
CC of the prostate. Prostate metastases are also known to secrete
CC immunologically reactive PSA since serum PSA is detectable at high levels
CC in prostatectomised patients showing widespread metastatic prostate
CC cancer. Cytotoxic compounds that could be activated by the proteolytic
CC activity of PSA should also be prostate cell specific as well as specific
CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
CC and this oligopeptide are useful in treatment of prostate cancer. The
CC oligopeptide can also be used in a quantitative assay for enzymatically
CC active PSA.
CC Sequence 9 AA;
Query Match 87.5%; Score 42; DB 26; Length 9;

```

Best Local Similarity 100.0%; Pred. No. 1.01e+02; Mismatches 0; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 nkisyq 7
 |||||
 QY 2 NKISYQ 7

RESULT 13
 ID W33488 standard; Peptide; 9 AA.
 AC W33488;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 68 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 7..8 /note= "PSA specific cleavage site"
 FT W09712624-A1.
 PN 10-APR-1997.
 PD 02-OCT-1996; U15713.
 PF 06-OCT-1995; US-540412.
 PR (MERI) MERCK & CO INC.
 PA DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 PI WPI; 97-225974/20.
 DR Oligopeptide(s) recognised and cleaved by free prostate specific antigen.
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 6; Page 102; 192pp; English.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 10 AA;

Query Match 87.5%; Score 42; DB 26; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 nkisyq 7
 |||||
 QY 2 NKISYQ 7

RESULT 14
 ID W33448 standard; Peptide; 10 AA.
 AC W33448;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 28 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.

OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 6..7 /note= "PSA specific cleavage site"
 FT W09712624-A1.
 PN 10-APR-1997.
 PD 02-OCT-1996; U15713.
 PF 06-OCT-1995; US-540412.
 PR (MERI) MERCK & CO INC.
 PA DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 PI WPI; 97-225974/20.
 DR Oligopeptide(s) recognised and cleaved by free prostate specific antigen.
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 6; Page 162; 192pp; English.
 CC The present sequence is a novel oligopeptide that is recognised and
 CC proteolytically cleaved by free prostate specific antigen (PSA). It was
 CC designed based on the sequences surrounding the PSA cleavage sites of
 CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
 CC chymotrypsin-like specificity is responsible for dissolution of the gel
 CC structure formed at ejaculation by proteolysis of the major gel proteins
 CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
 CC releases progressively motile spermatozoa. PSA complexed to
 CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
 CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
 CC of the prostate. Prostate metastases are also known to secrete
 CC immunologically reactive PSA since serum PSA is detectable at high levels
 CC in prostatectomised patients showing widespread metastatic prostate
 CC cancer. Cytotoxic compounds that could be activated by the proteolytic
 CC activity of PSA should also be prostate cell specific as well as specific
 CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
 CC and this oligopeptide are useful in treatment of prostate cancer. The
 CC oligopeptide can also be used in a quantitative assay for enzymatically
 CC active PSA.
 SQ Sequence 10 AA;

Query Match 87.5%; Score 42; DB 26; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
 |||||
 QY 2 NKISYQ 7

RESULT 15
 ID W33429 standard; Peptide; 10 AA.
 AC W33429;
 DT 27-MAR-1998 (first entry)
 DE Oligopeptide 9 cleaved by free prostate specific antigen.
 KW Oligopeptide; proteolytic cleavage; prostate specific antigen; PSA;
 KW cleavage site; Semenogelin I; chymotrypsin-like; serum PSA;
 KW adenocarcinoma; prostate metastases; prostate cancer; treatment;
 KW detection; cytotoxic conjugate; activation; quantitative assay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 5..6 /note= "PSA specific cleavage site"
 FT Modified_site 10 /note= "this Thr is optionally attached via a peptide bond to the 3'-amino group on the sugar ring of doxorubicin"
 FT W09712624-A1.
 PN 10-APR-1997.
 PD 02-OCT-1996; U15713.
 PF 06-OCT-1995; US-540412.
 PR (MERI) MERCK & CO INC.
 PA DeFeo-Jones D, Feng D, Garsky VM, Jones RE, Oliff AI;
 PI WPI; 97-225974/20.
 DR Oligopeptide(s) recognised and cleaved by free prostate specific antigen.
 PT antigen - useful in assays for active antigen, and in
 PT oligopeptide-drug conjugates for prostatic cancer treatment
 PS Claim 4 and 17; Page 161; 192pp; English.

CC The present sequence is a novel oligopeptide that is recognised and
CC proteolytically cleaved by free prostate specific antigen (PSA). It was
CC designed based on the sequences surrounding the PSA cleavage sites of
CC Semenogelin I, a major sperm entrapping gel protein. PSA, which has
CC chymotrypsin-like specificity is responsible for dissolution of the gel
CC structure formed at ejaculation by proteolysis of the major gel proteins
CC (Semenogelin I and II, and fibronectin). Liquefaction of the ejaculate
CC releases progressively motile spermatozoa. PSA complexed to
CC alpha 1-antichymotrypsin is the predominant form of serum PSA. Serum
CC measurements of PSA are useful for monitoring treatment of adenocarcinoma
CC of the prostate. Prostate metastases are also known to secrete
CC immunologically reactive PSA since serum PSA is detectable at high levels
CC in prostatectomised patients showing widespread metastatic prostate
CC cancer. Cytotoxic compounds that could be activated by the proteolytic
CC activity of PSA should also be prostate cell specific as well as specific
CC for PSA secreting prostate metastases. Conjugates of cytotoxic compounds
CC and this oligopeptide are useful in treatment of prostate cancer. The
CC oligopeptide can also be used in a quantitative assay for enzymatically
CC active PSA. In a study of cleavage affinity, the TFA salt of this peptide
CC was digested with enzymatically active PSA. After four hours 62 per cent
CC of the peptide was cleaved.
SQ Sequence 10 AA;

Query Match 87.5%; Score 42; DB 26; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.01e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 nkisyq 6
| | | | |
Qy 2 NKISYQ 7

Search completed: Thu Oct 28 11:33:57 1999
Job time : 17 secs.

THIS PAGE BLANK (USPTO)

W P S R L

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:34:13 1999; MasPar time 2.93 Seconds
Tabular output not generated. 95.729 Million cell updates/sec

Title: >US-09-081-707-12
Description: (1-7) from US09081707.pap
Perfect Score: 48
Sequence: 1 ENKISVQ 7

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.537; Variance 21.176; scale 0.970

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	48	100.0	582 2 A43412 semenogelin II precursor 4.61e+01
2	45	93.8	287 2 J02220 28k surface antigen 1 2.60e+00
3	43	89.6	709 2 E64213 DNA topoisomerase (EC 7.90e+00
4	42	87.5	384 2 A32885 xylose operon repress 1.36e+01
5	42	87.5	462 1 WTHUB semenogelin I precursor 1.36e+01
6	42	87.5	553 2 I49834 HE65 protein - Autogr 1.36e+01
7	42	87.5	879 2 S73757 hypothetical protein 1.36e+01
8	42	87.5	1021 2 E84576 hypothetical protein 1.36e+01
9	41	85.4	312 1 S28006 RNA-directed DNA poly 2.31e+01
10	41	85.4	319 2 D64303 hypothetical protein 2.31e+01
11	41	85.4	439 2 J04036 phosphopyruvate hydra 2.31e+01
12	41	85.4	544 2 S75388 probable phenylalanin 2.31e+01
13	41	85.4	704 2 I47228 carbonic anhydrase II 2.31e+01
14	40	83.3	276 2 S97072 transcription antiter 3.90e+01
15	40	83.3	374 2 S75459 hypothetical protein 3.90e+01
16	40	83.3	967 2 A64710 type III restriction 3.90e+01
17	40	83.3	969 2 I71810 type III restriction 3.90e+01
18	40	83.3	1108 2 I59385 guanylate cyclase (EC 3.90e+01
19	40	83.3	2958 2 S64921 probable membrane pro 3.90e+01
20	39	81.3	144 2 A61158 aminoglycoside N6'-ac 6.51e+01
21	39	81.3	187 2 G64323 ribosomal protein S4 6.51e+01
22	39	81.3	287 2 G69141 hypothetical protein 6.51e+01
23	39	81.3	344 2 A42359 fibrial adhesin F17- 6.51e+01

ALIGNMENTS									
RESULT	1								
ENTRY	A43412	#type complete							
TITLE	semenogelin II precursor - human								
ORGANISM	#formal_name Homo sapiens #common_name man								
DATE	30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-May-1998								
ACCESSIONS	A43412; B31489; A45295; S29156; S68765; S68762								
REFERENCE	A43412								
#authors	Ulvback, M.; Lazure, C.; Lilja, H.; Spurr, N.K.; Rao, V.V.; Loffler, C.; Hansmann, I.; Lundwall, A.								
#journal	J. Biol. Chem. (1992) 267:18080-18084								
#title	Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20.								
#cross-references	MUID:92388176								
#accession	A43412								
#molecule_type	DNA								
#residues	1-582 #label ULV								
#cross-references	GB:M81651; NID:G307417; PID:G307418								
#note	sequence extracted from NCBI backbone (NCBIN:112887, NCBIPI:112889)								
REFERENCE	A31489								
#authors	Lilja, H.; Abrahamsson, P.A.; Lundwall, A.								
#journal	J. Biol. Chem. (1989) 264:1894-1900								
#title	Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa.								
#cross-references	MUID:89109215								
#accession	B31489								
#status	nucleic acid sequence not shown								
#molecule_type	mRNA								
#residues	214-278, 'Y', 280-281 #label LIL								
REFERENCE	A45295								
#authors	Lilja, H.; Lundwall, A.								
#journal	Proc. Natl. Acad. Sci. U.S.A. (1992) 89:4559-4563								
#title	Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein.								
#cross-references	MUID:92262479								
#accession	A45295								
#molecule_type	mRNA								
#residues	3-582 #label LIL2								
#cross-references	GB:M81652								
REFERENCE	S29155								
#authors	Schneider, K.; Kausler, W.; Tripiet, D.; Jouvonal, K.; Spittler, G.								
#journal	Biol. Chem. Hoppe-Seyler (1989) 370:353-356								
#title	Isolation and structure determination of two peptides								

```

occurring in human seminal plasma.
#accession S29156
#molecule_type protein
#residues 390,'E',392-396,'E',398-400,'EW',403-405,'C',407,'DE',
#label SCH
#note this report is of a secondary sequence determined
simultaneously with the sequence with accession number
S29155 (see entry WIH0B); as a secondary sequence, it
should be considered less than fully reliable

REFERENCE
#authors Kise, H.; Nishioka, J.; Kawamura, J.; Suzuki, K.
#journal Eur. J. Biochem. (1996) 238:88-96
#title Characterization of semenogelin II and its molecular
interaction with prostate-specific antigen and protein C
inhibitor.
#cross-references MUID:96248425
#accession S68765
#molecule_type protein
#residues 105-107,'K',109-111;113-122;260-269;280-283 #label KIS
REFERENCE
#authors Malm, J.; Hellman, J.; Magnusson, H.; Laurell, C.B.; Lilja,
H.
#journal Eur. J. Biochem. (1996) 238:48-53
#title Isolation and characterization of the major gel proteins in
human semen, semenogelin I and semenogelin II.
#accession S68762
#molecule_type protein
#residues 420-421,'G',423-423 #label MAL
GENETICS
#gene GDB:SEMG2
#cross-references GDB:L32657; OMIM:182141
#map_position 20q12-20q13.1
#introns 26/1
CLASSIFICATION #superfamily semenogelin
KEYWORDS duplication; glycoprotein; semen; seminal vesicle; tandem
repeat
FEATURE
1-23 #domain signal sequence #status predicted #label SIG\
24-582 #product semenogelin II #status predicted #label MAT\
272 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 582 #molecular-weight 65444 #checksum 9704
Query Match 100.0%; Score 48; DB 2: Length 582;
Best Local Similarity 100.0%; Pred. No. 4.61e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 343 ENKISYQ 349
|||||
QY 1 ENKISYQ 7

RESULT 2
ENTRY JE0220 #type complete
TITLE 28k surface antigen 1 - Ehrlichia canis
ORGANISM #formal_name Ehrlichia canis
DATE 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change
17-Mar-1999
ACCESSIONS JE0220
REFERENCE JE0216
#authors Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.;
Burridge, M.J.; Allenan, A.R. (1998) 247:636-643
#journal Biochem. Biophys. Res. Commun. (1998) 247:636-643
#title Molecular characterization of a 28kDa surface antigen gene
family of the tribe Ehrlichiae.
#cross-references MUID:98321180
#accession JE0220
#molecule_type DNA
#residues 1-287 #label RED
#cross-references GB:AF062762
SUMMARY #length 287 #molecular-weight 32014 #checksum 463
Query Match 93.8%; Score 45; DB 2: Length 287;

occurring in human seminal plasma.
#accession S29156
#molecule_type protein
#residues 390,'E',392-396,'E',398-400,'EW',403-405,'C',407,'DE',
#label SCH
#note this report is of a secondary sequence determined
simultaneously with the sequence with accession number
S29155 (see entry WIH0B); as a secondary sequence, it
should be considered less than fully reliable

REFERENCE
#authors Kise, H.; Nishioka, J.; Kawamura, J.; Suzuki, K.
#journal Eur. J. Biochem. (1996) 238:88-96
#title Characterization of semenogelin II and its molecular
interaction with prostate-specific antigen and protein C
inhibitor.
#cross-references MUID:96248425
#accession S68765
#molecule_type protein
#residues 105-107,'K',109-111;113-122;260-269;280-283 #label KIS
REFERENCE
#authors Malm, J.; Hellman, J.; Magnusson, H.; Laurell, C.B.; Lilja,
H.
#journal Eur. J. Biochem. (1996) 238:48-53
#title Isolation and characterization of the major gel proteins in
human semen, semenogelin I and semenogelin II.
#accession S68762
#molecule_type protein
#residues 420-421,'G',423-423 #label MAL
GENETICS
#gene GDB:SEMG2
#cross-references GDB:L32657; OMIM:182141
#map_position 20q12-20q13.1
#introns 26/1
CLASSIFICATION #superfamily semenogelin
KEYWORDS duplication; glycoprotein; semen; seminal vesicle; tandem
repeat
FEATURE
1-23 #domain signal sequence #status predicted #label SIG\
24-582 #product semenogelin II #status predicted #label MAT\
272 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 582 #molecular-weight 65444 #checksum 9704
Query Match 100.0%; Score 48; DB 2: Length 582;
Best Local Similarity 100.0%; Pred. No. 4.61e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 343 ENKISYQ 349
|||||
QY 1 ENKISYQ 7

RESULT 2
ENTRY JE0220 #type complete
TITLE 28k surface antigen 1 - Ehrlichia canis
ORGANISM #formal_name Ehrlichia canis
DATE 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change
17-Mar-1999
ACCESSIONS JE0220
REFERENCE JE0216
#authors Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.;
Burridge, M.J.; Allenan, A.R. (1998) 247:636-643
#journal Biochem. Biophys. Res. Commun. (1998) 247:636-643
#title Molecular characterization of a 28kDa surface antigen gene
family of the tribe Ehrlichiae.
#cross-references MUID:98321180
#accession JE0220
#molecule_type DNA
#residues 1-287 #label RED
#cross-references GB:AF062762
SUMMARY #length 287 #molecular-weight 32014 #checksum 463
Query Match 93.8%; Score 45; DB 2: Length 287;

Best Local Similarity 85.7%; Pred. No. 2.60e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 214 QNKISYQ 220
|||||
QY 1 ENKISYQ 7

RESULT 3
ENTRY E64213 #type complete
TITLE DNA topoisomerase (EC 5.99.1.2) - Mycoplasma genitalium
(SGC3)
ALTERNATE_NAMES type I DNA topoisomerase
ORGANISM #formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
10-Oct-1997
ACCESSIONS E64213
REFERENCE A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession E64213
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-709 #label TIGR
#cross-references GB:U39691; GB:L43967; NID:g1045794; PID:g1045802;
TIGR:MG122
#experimental_source strain G-37
GENETICS
#genetic_code SGC3
KEYWORDS isomerase
SUMMARY #length 709 #molecular-weight 82544 #checksum 1986
Query Match 89.6%; Score 43; DB 2: Length 709;
Best Local Similarity 71.4%; Pred. No. 7.90e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 577 ENKVNQY 583
|||||
QY 1 ENKISYQ 7

RESULT 4
ENTRY A32885 #type complete
TITLE xylose operon repressor protein xylR - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 22-Nov-1989 #sequence_revision 22-Nov-1989 #text_change
16-Oct-1998
ACCESSIONS A32885; E69735
REFERENCE A32885
#authors Kreuzer, P.; Gaertner, D.; Allmansberger, R.; Hillen, W.
J. Bacteriol. (1989) 171:3840-3845
#journal Identification and sequence analysis of the Bacillus subtilis
W23 xylR gene and xyl operator.
#cross-references MUID:89291732
#accession A32885
#molecule_type DNA
#residues 1-384 #label KRE
#cross-references GB:M27248; NID:g143840; PID:g143841
#experimental_source strain W23
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,

```

A.: Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, V.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conner, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medique, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogilwa, A.; Oueda, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
#cross-references MUID:98044033
#accession E69735
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 35-53, 'SM', 56-82, 'V', 84-92, 'N', 94-103, 'V', 105-107, 'YR', 110-132, 'Q', 134-140, 'F', 142-154, 'D', 156-173, 'S', 175-178, 'Y', 180-181, 'S', 183-197, 'L', 199-205, 'D', 207-210, 'V', 212-277, 'L', 279-284, 'N', 286-320, 'V', 322-337, 'M', 339-347, 'S', 349-362, 'Q', 364-375, 'D', 377-380, 'Mn', 384 #label KUN
#cross-references GB:D99113; GB:AL009126; NID:g2634090; PID:eil83418; PID:g2634143
#experimental_source strain 168

GENETICS
#gene xylR
#start_codon GTG
CLASSIFICATION #superfamily xylose repressor; glucose kinase homology
KEYWORDS DNA binding; transcription regulation
FEATURE #domain glucose kinase homology #label GRH
141-269
SUMMARY #length 384 #molecular_weight 42295 #checksum 6179

Query Match 87.5%; Score 42; DB 2; Length 384;
Best Local Similarity 71.4%; Pred. No. 1.36e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 275 EKKVSYQ 281
|:|:|:|
QY 1 ENKISYQ 7

RESULT 5
ENTRY WTHUB #type complete
TITLE semenogelin I precursor - human
CONTAINS #molecule_type protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 04-Dec-1986 #sequence_revision 03-Oct-1995 #text_change 26-Feb-1999

ACCESSIONS BA3412; A31489; A91335; S29155; A43500; A91320; S29380; S68761; A03254
REFERENCE A43412
#authors Ulvback, M.; Lazure, C.; Lilja, H.; Spurr, N.K.; Rao, V.V.; Lofler, C.; Hansmann, I.; Lundwall, A.
#journal J. Biol. Chem. (1992) 267:18080-18084
#title Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20.
#cross-references MUID:92388176
#accession B43412
#molecule_type DNA
#residues 1-462 #label ULV
#cross-references GB:M81650; NID:g307416; PID:g487420
REFERENCE A31489
#authors Lilja, H.; Abrahamsson, P.A.; Lundwall, A.
#journal J. Biol. Chem. (1989) 264:1894-1900
#title Semenogelin, the predominant protein in human semen. Primary structure and identification of closely related proteins in the male accessory sex glands and on the spermatozoa.
#cross-references MUID:89109215
#accession A31489
#molecule_type mRNA
#residues 1-78, 'T', 80-422, 'K', 424-462 #label LIL
#cross-references GB:J04440
#note Ser-79 was also found
REFERENCE A91335
#authors Lilja, H.; Jeppsson, J.O.
#journal FEBS Lett. (1985) 182:181-184
#title Amino acid sequence of the predominant basic protein in human seminal plasma.
#cross-references MUID:85127550
#accession A91335
#molecule_type protein
#residues 108-159 #label LIL
#note this sequence represents a naturally occurring fragment from proteolytic cleavage of semenogelin during liquefaction of semen
REFERENCE S29155
#authors Schneider, K.; Kausler, W.; Tripler, D.; Jouvenal, K.; Spittler, G.
#journal Biol. Chem. Hoppe-Seyler (1989) 370:353-356
#title Isolation and structure determination of two peptides occurring in human seminal plasma.
#accession S29155
#molecule_type protein
#residues 316-320, 'L', 322-344 #label SCH
#note this sequence represents the amino end of a naturally occurring fragment from proteolytic cleavage of semenogelin during liquefaction of semen
REFERENCE A43500
#authors Ramasharma, K.; Sairam, M.R.; Seidah, N.G.; Chretien, M.; Manjunath, P.; Schiller, P.W.; Yamashiro, D.; Li, C.H.
#journal Science (1984) 223:1199-1202
#title Isolation, structure, and synthesis of a human seminal plasma peptide with inhibin-like activity.
#cross-references MUID:84146751
#accession A43500
#molecule_type protein
#residues 108-138 #label RAM
#note this sequence represents a naturally occurring fragment from proteolytic cleavage of semenogelin during liquefaction of semen
REFERENCE A91320
#authors Seidah, N.G.; Ramasharma, K.; Sairam, M.R.; Chretien, M.
#journal FEBS Lett. (1984) 167:98-102
#title Partial amino acid sequence of a human seminal plasma peptide with inhibin-like activity.
#cross-references MUID:84132557
#accession A91320
#molecule_type protein
#residues 108-138 #label SEI
#note this sequence represents a naturally occurring fragment


```

#authors      Tomb, J.F.; White, O.; Kervlavage, A.R.; Clayton, R.A.;
              Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
              H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
              J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
              Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.;
              McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
              Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
              Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
              Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes,
              W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
              C.M.; Venter, J.C.

#journal      Nature (1997) 388:539-547
#title        The complete genome sequence of the gastric pathogen
              Helicobacter pylori.
#cross-references MUID:97394467
#accession     E54576
#status        preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues      1-1021 ##label TOM
#cross-references GB:AE000560; GB:AE000511; NID:g2313554; PID:g2313560;
              TIGR:HP0453
#length 1021 #molecular-weight 120112 #checksum 8085
SUMMARY
Query Match      87.5%; Score 42; DB 2; Length 1021;
Best Local Similarity 71.4%; Pred. No. 1.36e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 558 EKIA9Q 564
|:|:|:|:|
Qy 1 ENKISYQ 7

RESULT 9
ENTRY S28006 #type complete
TITLE RNA-directed DNA polymerase (EC 2.7.7.49), msDNA specific -
      Escherichia coli retron Ec79
ALTERNATE_NAMES DNA nucleotidyltransferase (RNA-directed); reverse
      transcriptase; revertase
ORGANISM #formal_name Escherichia coli retron Ec79
DATE 17-Apr-1993 #sequence_revision 15-Oct-1996 #text_change
      05-Sep-1997
ACCESSIONS S28006; S22963
REFERENCE Lim, D.
#authors Mol. Microbiol. (1992) 6:3531-3542
#journal Structure and biosynthesis of unbranched multicopy
#title single-stranded DNA by reverse transcriptase in a clinical
      Escherichia coli isolate.
#cross-references MUID:93116591
#accession S28006
#molecule_type DNA
#residues 1-312 ##label LIM
#cross-references EMBL:212832; NID:g42499; PID:g42501
#experimental_source E. coli strain 161
GENETICS
#note insertion site is equivalent to 6 min of E. coli K12 genetic
      map
CLASSIFICATION #superfamily reverse transcriptase.
KEYWORDS nucleotidyltransferase
SUMMARY #length 312 #molecular-weight 35734 #checksum 2668

Query Match      85.4%; Score 41; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.31e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 175 ENKISY 180
|:|:|:|:|
Qy 1 ENKISY 6

RESULT 10
ENTRY D64303 #type complete

```

```

TITLE          hypothetical protein HI1305 homolog - Methanococcus
              jannaschii
ORGANISM        #formal_name Methanococcus jannaschii
DATE            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
              10-Oct-1997
ACCESSIONS      D64303
REFERENCE        A64300
#authors        Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
              R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
              R.A.; Gocayne, J.D.; Kervlavage, A.R.; Dougherty, B.A.;
              Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
              Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
              Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
              J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
              J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
              K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
              H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal        Science (1996) 273:1058-1073
#title          Complete genome sequence of the methanogenic archaeon,
              Methanococcus jannaschii.
#cross-references MUID:96337999
#accession      D64303
#status        preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues      1-319 ##label BUL
#cross-references GB:U67461; GB:L77117; NID:gl590827; PID:gl590830;
              TIGR:MJ0028; PID:gl510201
GENETICS
#map_position REV29473-28514
SUMMARY #length 319 #molecular-weight 36160 #checksum 7290

Query Match      85.4%; Score 41; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 2.31e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 ENKISY 179
|:|:|:|:|
Qy 1 ENKISY 6

RESULT 11
ENTRY JC4036 #type complete
TITLE phosphopyruvate hydratase (EC 4.2.1.11) - fission yeast
      (Schizosaccharomyces pombe)
ALTERNATE_NAMES 2-phospho-D-glycerate hydrolyase; enolase
ORGANISM #formal_name Schizosaccharomyces pombe
DATE 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change
      08-Sep-1997
ACCESSIONS JC4036
REFERENCE Jackson, J.C.; Lopes, J.M.
#authors Gene (1995) 154:109-113
#journal A cDNA from Schizosaccharomyces pombe encoding a putative
#title enolase.
#cross-references MUID:95172389
#accession JC4036
#molecule_type mRNA
#residues 1-439 ##label JAC
#cross-references GB:U13799; NID:g535441; PID:g535442
#note The authors translated the codon GTC for residue 20 as B
      and GAC for residue 83 as Val
      and GAC for residue 83 as Val
      2-phosphoglycerate to phosphoenolpyruvate.
COMMENT This glycolytic enzyme catalyzes the dehydration of
      2-phosphoglycerate to phosphoenolpyruvate.
GENETICS
#gene enol
CLASSIFICATION #superfamily enolase
KEYWORDS carbon-oxygen lyase; gluconeogenesis; glycolysis; hydro-lyase
SUMMARY #length 439 #molecular-weight 47323 #checksum 4121

Query Match      85.4%; Score 41; DB 2; Length 439;
Best Local Similarity 71.4%; Pred. No. 2.31e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 270 ENKITYQ 276
  |||:|
QY 1 ENKISYQ 7

RESULT 12
ENTRY #type complete
TITLE Probable phenylalanine-tRNA ligase (EC 6.1.1.20) beta chain
ALTERNATE_NAMES - Sulfolobus solfataricus
ORGANISM phenylalanyl-tRNA synthetase beta chain; protein c04021
DATE #formal_name Sulfolobus solfataricus
09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change
17-Mar-1999

ACCESSIONS S75388
REFERENCE S73076
#authors Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.L.; Young, F.; Schenk, M.E.; Gaasterland, T.; Doolittle, W.F.; Ragan, M.A.; Charlebois, R.L.
#journal Mol. Microbiol. (1996) 22:175-191
#title Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from Sulfolobus solfataricus P2.
#cross-references M01D:97055432.
#accession S75388
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-544 ##label SEN
##cross-references EMBL:Y08257; NID:g1707772; PID:e283832; PID:g1707793
##experimental_source strain P2
##note the nucleotide sequence was submitted to the EMBL Data Library, September 1996

CLASSIFICATION #superfamily yeast cytosolic phenylalanine-tRNA ligase alpha chain
KEYWORDS aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
SUMMARY #length 544 #molecular-weight 61815 #checksum 7376

Query Match 85.4%; Score 41; DB 2; Length 544;
Best Local Similarity 57.1%; Pred. No. 2.31e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 459 DNKYSYE 465
  |||:|
QY 1 ENKISYQ 7

RESULT 13
ENTRY #type complete
TITLE carbonic anhydrase II inhibitor (transferrin homolog) precursor - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change
15-Jan-1999
ACCESSIONS I47228
REFERENCE I47228
#authors Roush, E.D.; Fierke, C.A.
#journal Biochemistry (1992) 31:12536-12542
#title Purification and characterization of a carbonic anhydrase II inhibitor from porcine plasma.
#cross-references M01D:93099129
#accession I47228
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-704 ##label ROU
##cross-references EMBL:U36916; NID:g1016329; PID:g1016330

GENETICS
#gene pICA
CLASSIFICATION #superfamily transferrin; transferrin repeat homology
KEYWORDS duplication
FEATURES
#domain transferrin repeat homology #label TRH1
SUMMARY #length 704 #molecular-weight 77634 #checksum 8035

```

```

Query Match 85.4%; Score 41; DB 2; Length 704;
Best Local Similarity 57.1%; Pred. No. 2.31e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 667 QDKINYQ 673
  |||:|
QY 1 ENKISYQ 7

RESULT 14
ENTRY #type complete
TITLE transcription antiterminator of sacA/sacP sacT - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 07-Oct-1994 #sequence_revision 22-Nov-1996 #text_change
17-Mar-1999
ACCESSIONS S39702; C69703
REFERENCE S39655
#authors Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu, M.; Lubochinsky, B.; Marcelino, L.; Moszer, I.; Presecan, E.; Santana, M.; Schneider, E.; Schweizer, J.; Vertes, A.; Rapoport, G.; Danchin, A.
#journal Mol. Microbiol. (1993) 10:371-384
#title Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees.
#cross-references M01D:95020537
#accession S39702
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-276 ##label GLA
##cross-references EMBL:X73124; NID:g413923; PID:g580872
##note the nucleotide sequence was submitted to the EMBL Data Library, June 1993

REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerf-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mael, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

```

#cross-references MUID:98044033
#accession C69703
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-276 ##label KUN
##cross-references GB:299123; GB:AL009126; NID:g2636240; PID:el186306;
PID:g2636342
##experimental_source strain 168

GENETICS

#gene sacT
#start_codon TTG.
CLASSIFICATION #superfamily Bacillus subtilis transcription antiterminator
licT

SUMMARY #length 276 #molecular-weight 32074 #checksum 7428

Query Match 83.3%; Score 40; DB 2; Length 276;
Best Local Similarity 85.7%; Pred. No. 3.90e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 194 ENSISYQ 200

|||||

QY 1 ENKISYQ 7

RESULT 15
ENTRY S75459 #type complete
TITLE hypothetical protein slf1512 - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998

ACCESSIONS S75459
REFERENCE S74322

#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

#cross-references MUID:97061201
#accession S75459
##status preliminary
##molecule_type DNA
##residues 1-374 ##label KAN
##cross-references EMBL:D90911; GB:AB001339; NID:g1653083; PID:dl018753;
PID:g1653104

##note the nucleotide sequence was submitted to the EMBL Data Library, June 1996.

SUMMARY #length 374 #molecular-weight 39647 #checksum 6514

Query Match 83.3%; Score 40; DB 2; Length 374;
Best Local Similarity 71.4%; Pred. No. 3.90e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 127 ESKISYE 133

|||||

QY 1 ENKISYQ 7

Search completed: Thu Oct 28 11:34:26 1999
Job time : 13 secs.

THIS PAGE BLANK (USPTO)

W A S E R H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Oct 28 11:34:43 1999; MasPar time 2.06 Seconds

Tabular output not generated. 95.967 Million cell updates/sec

Title: >US-09-081-707-12
Description: (1-7) from US09081707.ppe
Perfect Score: 48
Sequence: 1 ENKISQ 7
Scoring table: PAM 150
Gap 15
Searched: 77977 seqs, 28268293 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 21.081; Variance 18.718; scale 1.126

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	100.0	582	1 SEM2_HUMAN	SEMENOGELIN II PRECURS	9.82e-02
2	48	100.0	706	1 SEM2_MACMU	SEMENOGELIN II PRECURS	9.82e-02
3	43	89.6	709	1 TOPI_MYCGE	DNA TOPOISOMERASE I (E	2.43e+00
4	42	87.5	384	1 XYL1_BACSU	XYLOSE REPRESSOR.	4.47e+00
5	42	87.5	462	1 SEM1_HUMAN	SEMENOGELIN I PROTEIN	4.47e+00
6	42	87.5	553	1 VH65_NPVAC	EARLY 65 KD PROTEIN.	4.47e+00
7	41	85.4	319	1 Y028_METJA	HYPOTHETICAL PROTEIN M	8.13e+00
8	41	85.4	439	1 ENOL_SCHPO	ENOLASE (EC 4.2.1.11)	8.13e+00
9	40	83.3	276	1 SACT_BACSU	SACPA OPERON ANTITERMI	1.46e+01
10	40	83.3	1108	1 CYGF_HUMAN	RETINAL GUANYLYL CYCLA	1.46e+01
11	39	81.3	187	1 RS4_METJA	30S RIBOSOMAL PROTEIN	2.59e+01
12	39	81.3	381	1 CYB_NOTPE	CYTOCHROME B.	2.59e+01
13	39	81.3	434	1 PYRP_BACSU	URACIL PERMEASE (URACI	2.59e+01
14	39	81.3	458	1 IGBH_ECOLI	HYPOTHETICAL 52.8 KD P	2.59e+01
15	39	81.3	609	1 TH13_YEAST	THIAMINE METABOLISM RE	2.59e+01
16	39	81.3	625	1 BAR3_SCHCO	PEROMONE B ALPHA 3 RE	2.59e+01
17	39	81.3	639	1 BAR1_SCHCO	PEROMONE B ALPHA 1 RE	2.59e+01
18	39	81.3	689	1 YE30_HELPY	HYPOTHETICAL PROTEIN H	2.59e+01
19	39	81.3	1087	1 PG85_XENLA	ALPHA PLATELET-DERIVED	2.59e+01
20	39	81.3	1089	1 NM02_YEAST	NONSENSE-MEDIATED MRNA	2.59e+01
21	39	81.3	1820	1 CINA_ELELE	SODIUM CHANNEL PROTEIN	2.59e+01
22	38	79.2	94	1 YQGV_BACSU	HYPOTHETICAL 10.5 KD P	4.53e+01
23	38	79.2	145	1 Y127_MYCGE	HYPOTHETICAL PROTEIN M	4.53e+01

24	38	79.2	163	1 TPX_STRPA	PROBABLE THIOL PEROXID	4.53e+01
25	38	79.2	191	1 CIMA_YEAST	GTP-BINDING PROTEIN CI	4.53e+01
26	38	79.2	349	1 Y4F_SCHPO	HYPOTHETICAL 40.9 KD P	4.53e+01
27	38	79.2	432	1 ENO_TREPA	ENOLASE (EC 4.2.1.11)	4.53e+01
28	38	79.2	455	1 BIOA_BACSH	ADENOSYLMETHIONINE-8-A	4.53e+01
29	38	79.2	487	1 ACH6_CAEEL	ACETYLCHOLINE RECEPTOR	4.53e+01
30	38	79.2	544	1 RGSE_RAT	REGULATOR OF G-PROTEIN	4.53e+01
31	38	79.2	545	1 PYRG_HAEN	CTP SYNTHASE (EC 6.3.4	4.53e+01
32	38	79.2	547	1 RGSE_MOUSE	REGULATOR OF G-PROTEIN	4.53e+01
33	38	79.2	579	1 UL25_VZVD	VIROIN GENE 34 PROTEIN	4.53e+01
34	38	79.2	604	1 VEI_BPV2	REPLICATION PROTEIN E1	4.53e+01
35	38	79.2	605	1 VEI_BPV1	REPLICATION PROTEIN E1	4.53e+01
36	38	79.2	612	1 GIDA_MYCPN	GLUCOSE INHIBITED DIVI	4.53e+01
37	38	79.2	613	1 VEI_PAPVD	REPLICATION PROTEIN E1	4.53e+01
38	38	79.2	656	1 DNAA_STRCO	CHROMOSOMAL REPLICATIO	4.53e+01
39	38	79.2	1257	1 FLIH_CAEEL	FLIGHTLESS-I PROTEIN H	4.53e+01
40	38	79.2	1738	1 YCF1_EPIVI	HYPOTHETICAL 208 KD PR	4.53e+01
41	37	77.1	191	1 Y948_METJA	HYPOTHETICAL PROTEIN M	7.83e+01
42	37	77.1	357	1 YEV7_YEAST	HYPOTHETICAL 40.8 KD P	7.83e+01
43	37	77.1	445	1 WAPA_STRMU	WALL-ASSOCIATED PROTEI	7.83e+01
44	37	77.1	728	1 YBJ7_YEAST	HYPOTHETICAL 83.0 KD P	7.83e+01
45	37	77.1	1019	1 VP3_RDV	MAJOR 114 KD STRUCTURA	7.83e+01

ALIGNMENTS

RESULT 1
ID SEM2_HUMAN STANDARD; PRT; 582 AA.
AC Q02383;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SEMENOGELIN II PRECURSOR (SGII).
GN SEMG2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SEMINAL VESICLE;
RX MEDLINE; 92262479.
RA LUNDWALL A., LILJA H.;
RT "Molecular cloning of epididymal and seminal vesicular transcripts encoding a semenogelin-related protein";
RT Proc. Natl. Acad. Sci. U.S.A. 89:4559-4563(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92388176.
RA ULYSBAECK M., LAZURE C., LILJA H., SPURR N.K., RAO V.V., LOEFFLER C., HANSMANN I., LUNDWALL A.;
RT "Gene structure of semenogelin I and II. The predominant proteins in human semen are encoded by two homologous genes on chromosome 20."; J. Biol. Chem. 267:18080-18084(1992).
CC -!- FUNCTION: THREE SEMENOGELIN PROTEINS ARE FOUND IN HUMAN SEMEN, OF WHICH THE 52 KD SEMENOGELIN I IS THE MOST ABUNDANT. THEY PARTICIPATE IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS OF SEMENOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE PROTEASE.
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEXES OF SEMENOGELIN I AND THE OTHER SEMENOGELIN PROTEINS, THE 71- AND 76-KD POLYPEPTIDES..
CC -!- TISSUE SPECIFICITY: SEMINAL VESICLES, AND TO A MUCH LESSER EXTENT, EPIDIDYMIS.
CC -!- PTM: SEMENOGELIN II IS THOUGHT TO FORM BOTH THE 71-KD POLYPEPTIDE AND, IN ITS GLYCOSYLATED FORM, THE 76-KD POLYPEPTIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
DR EMBL; M81651; G307418; -
DR EMBL; M81652; G38239; -
DR EMBL; Z47556; E82662; -
DR FIR; A43412; A43412.
DR MIM; I82141; -
KW SEMEN; SEMINAL VESICLE; REPEAT; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 23
FT CHAIN 24 582 SEMENOGELIN II.
FT DOMAIN 70 559 REPEAT-RICH REGION.
FT REPEAT 70 129 2-1.
FT REPEAT 141 200 2-2.
FT REPEAT 201 260 2-2.
FT DOMAIN 261 500 4 X 60 AA TANDEM REPEATS, TYPE I.
FT REPEAT 501 559 3-2.
FT CARBOHYD 272 272 PROBABLE.
SQ SEQUENCE 582 AA; 65445 MW; DD20304E CRC32;

Query Match 100.0%; Score 48; DB 1; Length 582;
Best Local Similarity 100.0%; Pred. No. 9.82e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 343 ENKISYQ 349
|||||
QY 1 ENKISYQ 7

RESULT 2
ID SEM2_MACMU STANDARD; PRT; 706 AA.
AC Q95196;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SEMENOGELIN II PRECURSOR (S011).
GN SEMG2.
OS MACACA MULATTA (RHESUS MACAQUE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPIITHECIDAE; CERCOPIITHECINAE; MACACA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97274635.
RA OLIVSBACK M., LUNDWALL A.;
RT "Cloning of the semenogelin II gene of the rhesus monkey.
RT Duplications of 360 bp extend the coding region in man, rhesus monkey
RT and baboon."
RL EUR. J. BIOCHEM. 245:25-31(1997).
CC -!- FUNCTION: PARTICIPATE IN THE FORMATION OF A GEL MATRIX ENTRAPPING
CC THE ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
-----
KW EMBL; X92589; E208370; -
KW SEMEN; SEMINAL VESICLE; REPEAT; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 706 SEMENOGELIN II.
SQ SEQUENCE 706 AA; 79875 MW; 9D5897A8 CRC32;
```

Query Match 100.0%; Score 48; DB 1; Length 706;
Best Local Similarity 100.0%; Pred. No. 9.82e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 347 ENKISYQ 353
|||||
QY 1 ENKISYQ 7

```
-----
RESULT 3
ID TOPL_MYCGE STANDARD; PRT; 709 AA.
AC P47368;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DNA TOPOISOMERASE I (EC 5.99.1.2) (OMEGA-PROTEIN) (RELAXING ENZYME)
DE (UNWISTING ENZYME) (SWIVELASE).
GN TOPA OR MG122.
OS MYCOPLASMA GENITALIUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITCHMANN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RT SCIENCE 270:397-403(1995).
RN [2]
RP SEQUENCE OF 399-481 AND 527-657 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE; 94075230.
RA PETERSON S.N., HU P.-C., BOTT K.F., HUTCHISON C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RL J. BACTERIOL. 175:7918-7930(1993).
CC -!- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -!- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA BACKBONE BOND, IT
CC SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN WHICH A TYROSYL OXYGEN
CC IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS AT ONE END OF THE
CC ENZYME-SEVERED DNA STRAND.
CC -!- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
-----
DR EMBL; U39691; G1045802; -
DR EMBL; U02134; G409912; -
DR EMBL; U02242; G407259; -
DR TIGR; MG122; -
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
DR PFAM; PF01131; Topoisom_bac; 1.
DR HSSP; P06612; 1ECL.
KW ISOMERASE; TOPOISOMERASE; DNA-BINDING.
FT ACT-SITE 334 334 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 709 AA; 82544 MW; 18063FDC CRC32;
```

Query Match 89.6%; Score 43; DB 1; Length 709;
Best Local Similarity 71.4%; Pred. No. 2.43e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 577 ENKVNQ 583
|||||
QY 1 ENKISYQ 7

```
RESULT 4
ID XYLR_BACSU STANDARD; PRT; 384 AA.
AC P16557;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE XYLOSE REPRESSOR.
GN XYLR.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W23;
RX MEDLINE; 89291732.
RA KREUZER P., GAERTNER D., ALLMANSBERGER R., HILLEN W.;
RT "Identification and sequence analysis of the Bacillus subtilis W23
xyr gene and xyl operator.";
RL J. BACTERIOL. 171:3840-3845(1989).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF XYLOSE-UTILIZING ENZYMES.
CC -!- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27248; G143841; -
DR PIR; A32885; A32885
DR PROSITE; PS01125; ROK; 1.
DR PFAM; PF00480; ROK; 1.
RW TRANSCRIPTION REGULATION; DNA-BINDING; REPRESSOR; XYLOSE METABOLISM.
FT DNA_BIND 29 48 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 384 AA; 42295 MW; FEALAE85 CRC32;
-----
Query Match 87.5%; Score 42; DB 1; Length 384;
Best Local Similarity 71.4%; Pred. No. 4.47e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 275 EKKYSYQ 281
|:|:|
QY 1 ENKISYQ 7
-----
RESULT 5
ID SEM1_HUMAN STANDARD; PRT; 462 AA.
AC P04279;
DT 20-MAR-1987 (REL. 04, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SEMENOGELIN I PROTEIN PRECURSOR (SGI) [CONTAINS: SEMINAL BASIC
PROTEIN; ALPHA-INHIBIN-92; ALPHA-INHIBIN-31].
GN SEMG1 OR SEMG
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 89109215.
RA LILJA H., ABRAHAMSSON P.-A., LUNDWALL A.;
RT "Semnodelin, the predominant protein in human semen. Primary
structure and identification of closely related proteins in the male
accessory sex glands and on the spermatozoa.";
RL J. BIOL. CHEM. 264:1894-1900(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE; 92388176.
RA ULVSBAECK M., LAZURE C., LILJA H., SPURR N.K., RAO V.V., LOEFFLER C.,
RA HANSMANN I., LUNDWALL A.;
```

```
RT "Gene structure of semenogelin I and II. The predominant proteins in
human semen are encoded by two homologous genes on chromosome 20.";
J. BIOL. CHEM. 267:18080-18084(1992).
RN [3]
RP SEQUENCE OF 108-159.
RX MEDLINE; 85127550.
RA LILJA H., JEPSSON J.-O.;
RT "Amino acid sequence of the predominant basic protein in human
seminal plasma.";
RL FEBS LETT. 182:181-184(1985).
RN [4]
RP SEQUENCE OF 108-138.
RX MEDLINE; 84132557.
RA SEIDAH N.G., RAMASHARMA K., SAIRAM M.R., CHRETIEN M.;
RT "Partial amino acid sequence of a human seminal plasma peptide with
inhibin-like activity.";
RL FEBS LETT. 167:98-102(1984).
RN [5]
RP SEQUENCE OF 68-159.
RX MEDLINE; 85216629.
RA LI C.H., HAMMONDS R.G., RAMASHARMA K., CHUNG D.;
RT "Human seminal alpha inhibins: isolation, characterization, and
structure.";
RL PROC. NATL. ACAD. SCI. U.S.A. 82:4041-4044(1985).
CC -!- FUNCTION: SEMENOGELIN I IS THE PREDOMINANT PROTEIN IN HUMAN SEMEN.
CC IT PARTICIPATES IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE
CC ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS
CC OF SEMENOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY
CC CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS
CC THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE
CC PROTEASE.
CC -!- FUNCTION: ALPHA-INHIBIN-92 AND ALPHA-INHIBIN-31, DERIVED FROM
CC THE PROTEOLYTIC DEGRADATION OF SEMENOGELIN, INHIBIT THE SECRETION
CC OF PITUITARY FOLLICLE-STIMULATING HORMONE.
CC -!- SUBUNIT: OCCURS IN DISULFIDE-LINKED COMPLEXES WHICH MAY ALSO
CC CONTAIN TWO LESS ABUNDANT 71- AND 76-KD SEMENOGELIN-RELATED
CC POLYPEPTIDES.
CC -!- TISSUE SPECIFICITY: SEMINAL VESICLE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04440; G338019; -
DR EMBL; Z47556; E133812; -
DR EMBL; M81650; G487420; -
DR PIR; A03254; WTHUB.
DR PIR; A31489; A31489.
DR MIM; 182140; -
KW SEMEN; SEMINAL VESICLE; REPEAT; SIGNAL; POLYMORPHISM.
FT SIGNAL 1 23
FT CHAIN 24 462 SEMENOGELIN I.
FT MOD_RES 24 24 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT PEPTIDE 68 159 ALPHA-INHIBIN-92.
FT PEPTIDE 108 138 ALPHA-INHIBIN-31.
FT PEPTIDE 108 159 SEMINAL BASIC PROTEIN.
FT DISULFID 239 239 INTERCHAIN.
FT REPEAT 174 215 42 AA REPEAT 1.
FT REPEAT 234 275 42 AA REPEAT 2.
FT REPEAT 282 339 58 AA REPEAT 1.
FT REPEAT 342 399 58 AA REPEAT 2.
FT REPEAT 414 455 42 AA REPEAT 3.
FT VARIANT 79 79 S -> T (LESS COMMON GENETIC VARIANT).
FT CONFLICT 423 423 K -> N (IN REF. 2).
SQ SEQUENCE 462 AA; 52131 MW; C87515C7 CRC32;
-----
Query Match 87.5%; Score 42; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 4.47e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 284 NKISYQ 289
QY 2 NKISYQ 7

RESULT 6
ID VH65_NPVAC STANDARD; PRT; 553 AA.
AC Q08339;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE EARLY 65 KD PROTEIN.
GN HE65.
OS AUTOGRAHA CALIFORNICA NUCLEAR POLYHEDROSIS VIRUS (ACMPV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYHEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E;
RX MEDLINE; 93381786.
RA BECKER D., KNEBEL-MORSBORG D.;
RT "Sequence and temporal appearance of the early transcribed baculovirus gene HE65."
RL J. VIROL. 67:5867-5872(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE; 94303173.
RA AYRES M.D., HOWARD S.C., KUZIO J., LOPEZ-FERRER M., POSSEE R.D.;
RT "The complete DNA sequence of Autographa californica nuclear polyhedrosis virus."
RL VIROLOGY 202:586-605(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73577; G313680; -
DR EMBL; L22858; G559174; -
DR FIR; S35872; S35872.
KW EARLY PROTEIN.
SQ SEQUENCE 553 AA; 65576 MW; 420D4413 CRC32;

Query Match 87.5%; Score 42; DB 1; Length 553;
Best Local Similarity 85.7%; Pred. No. 4.47e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 271 ENKIVQ 277
QY 1 ENKISQ 7

RESULT 7
ID Y028_METJA STANDARD; PRT; 319 AA.
AC Q60337;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MJ0028.
GN MJ0028.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,

RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GEORHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii."
RL SCIENCE 273:1058-1073(1996).
CC -!- SIMILARITY: WEAK, TO THE HYPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67461; G1590830; -
DR TIGR; MJ0028; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 319 AA; 36160 MW; 606CC6D4 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 8.13e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 ENKISY 179
QY 1 ENKISY 6

RESULT 8
ID ENOL_SCHPO STANDARD; PRT; 439 AA.
AC P40370;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE ENOLASE (EC 4.2.1.11) (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE).
GN ENOL.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95172389.
RA JACKSON J.C., LOPES J.M.;
RT "A cDNA from Schizosaccharomyces pombe encoding a putative enolase."
RL GENE 154:109-113(1995).
CC -!- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE + H(2)O.
CC -!- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING THE DIMER.
CC -!- PATHWAY: GLYCOLYSIS.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13799; G535442; -
DR PROSITE; PS00164; ENOLASE; 1.
DR PFAM; PF00113; enolase; 1.
```

DR HSP: P00924; INEL.
KW LYASE; GLYCOLYSIS; MAGNESIUM.
FT ACT_SITE 159 159 BY SIMILARITY.
FT METAL 246 246 MAGNESIUM (BY SIMILARITY).
FT METAL 295 295 MAGNESIUM (BY SIMILARITY).
FT METAL 320 320 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 439 AA; 47323 MW; 61A96F15 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 439;
Best Local Similarity 71.4%; Pred. No. 8.13e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 270 ENKLYQ 276
||:||||
QY 1 ENKISYQ 7

RESULT 9
ID SACT_BACSU STANDARD; PRT; 276 AA.
AC P26212;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE SACTA OPERON ANTITERMINATOR.
GN SACT OR IPA-47D.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 90299824.
RA DEBAROUILLE M., ARNAUD M., FOUET A., KLIER A., RAPOPORT G.;
RT "the sact gene regulating the sactA operon in Bacillus subtilis
RT shares strong homology with transcriptional antiterminators.";
RL J. BACTERIOL. 172:3966-3973(1990).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 95020537.
RA GLASER P., KUNST F., ARNAUD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONESCU M., LUBOCHINSKY B., MARCELINO L., MOSZER I.,
RA PRESECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DANCHIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL MOL. MICROBIOL. 10:371-384(1993).
CC -1- FUNCTION: MEDIATES POSITIVE REGULATION OF THE SACTA OPERON BY
CC FUNCTIONING AS AN ANTITERMINATOR FACTOR OF TRANSCRIPTION.
CC -1- PTM: PHOSPHORYLATED AND INACTIVATED BY SACT (EII-SCP) (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTIONAL ANTITERMINATOR BGLG
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL: J03006; G143488; ALT_INIT.
DR EMBL: X73124; G580872;
DR EMBL: Z99123; E1186306;
DR SUBTILIST; BG10593; SACT.
DR PROSITE; PS00654; ANTITERMINATORS_BGLG; 1.
DR PFAM; PF00874; BglG_antitermin; 2.
DR HSP; P15401; 1A0U.
DR TRANSCRIPTION REGULATION; ACTIVATOR; RNA-BINDING; PHOSPHORYLATION.
KW VARIANT 96 96 D -> Y (IN SACT30 MUTANT).
FT SEQUENCE 276 AA; 32074 MW; 4FFB1288 CRC32;
SQ SEQUENCE 276 AA; 32074 MW; 4FFB1288 CRC32;

Query Match 83.3%; Score 40; DB 1; Length 276;
Best Local Similarity 85.7%; Pred. No. 1.46e+01;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 194 ENSISYQ 200
||:||||
QY 1 ENKISYQ 7

RESULT 10
ID CYGF_HUMAN STANDARD; PRT; 1108 AA.
AC P51841;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE RETINAL GUANYLYL CYCLASE 2 PRECURSOR (EC 4.6.1.2) (GUANYLATE CYCLASE
DE 2F, RETINAL) (RETGC-2) (ROD OUTER SEGMENT MEMBRANE GUANYLATE CYCLASE
DE 2) (ROS-GC2) (GUANYLATE CYCLASE F) (GC-F).
GN GUCY2F OR GUC2F OR RETGC2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-RETINA;
RX MEDLINE; 95296345.
RA LOWE D.G., DIZHOOR A.M., LIU K., GU Q., SPENCER M., LAURA R.,
RA LU L., HURLEY J.B.;
RT "Cloning and expression of a second photoreceptor-specific membrane
RT retina guanylyl cyclase (RetGC), RetGC-2,"
RL PROC. NATL. ACAD. SCI. U.S.A. 92:5535-5539(1995).
CC -1- FUNCTION: PROBABLY PLAYS A SPECIFIC FUNCTIONAL ROLE IN THE RODS
CC AND/OR CONES OF PHOTORECEPTORS. IT MAY BE THE ENZYME INVOLVED IN
CC THE RESYNTHESIS OF CGMP REQUIRED FOR RECOVERY OF THE DARK STATE
CC AFTER PHOTOTRANSDUCTION.
CC -1- CATALYTIC ACTIVITY: GTP -> 3',5'-CYCLIC GMP + PYROPHOSPHATE.
CC -1- ENZYME REGULATION: ACTIVATED BY GCAP-1; INHIBITED BY CALCIUM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: RETINA. LOCALIZED EXCLUSIVELY IN THE OUTER
CC NUCLEAR LAYER AND INNER SEGMENTS OF THE ROD AND CONE PHOTORECEPTOR
CC CELLS.
CC -1- PTM: THERE ARE 9 CONSERVED CYSTEINE RESIDUES IN SENSORY GUANYLATE
CC CYCLASES, 6 IN THE EXTRACELLULAR DOMAIN, WHICH MAY BE INVOLVED IN
CC INTRA- OR INTERCHAIN DISULFIDE BONDS.
CC -1- SIMILARITY: TO OTHER GUANYLATE CYCLASES. BELONGS TO THE SUBFAMILY
CC OF SENSORY GUANYLATE CYCLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L37378; G945225;
DR EMBL: L37378; G945225;
DR MIM; 300041;
DR PROSITE; PS00452; GUANYLATE_CYCLASES; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR PFAM; PF00211; guanylate_cyc; 1.
DR PFAM; PF01094; ANF_receptor; 1.
DR HSP; Q02846; 1A0L.
DR LYASE; CGMP SYNTHESIS; SIGNAL; TRANSMEMBRANE; MULTIGENE FAMILY;
KW VISION.
FT SIGNAL 1 50
FT CHAIN 51 1108
FT DOMAIN 51 467
FT TRANSMEM 468 490
FT DOMAIN 491 1108
FT DOMAIN 524 810
FT DOMAIN 813 1064
FT DISULFID 104 132
FT DISULFID 452 452
FT DISULFID 460 460
FT POTENTIAL.
FT RETINAL GUANYLYL CYCLASE 2.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN KINASE-LIKE.
FT CATALYTIC.
FT BY SIMILARITY.
FT INTERCHAIN (BY SIMILARITY).
FT INTERCHAIN (BY SIMILARITY).

```

SQ SEQUENCE 1108 AA: 124821 MW; F119341A CRC32;
Query Match 83.3%; Score 40; DB 1; Length 1108;
Best Local Similarity 83.3%; Pred. No. 1.46e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 157 DNKISY 162
:|||||
QY 1 ENKISY 6

RESULT 11
ID RS4_METUA STANDARD; PRT; 187 AA.
AC P54020;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 30S RIBOSOMAL PROTEIN S4P.
GN MJ0190.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY E.F., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GÖRGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KATNE B.P., BORODOVSKY M.,
RA KLEIN H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL SCIENCE 273:1058-1073(1996).
CC -!- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67475; G1498964; -.
DR TIGR; MT0190; -.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PFAM; PF00163; S4; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 187 AA: 22050 MW; CFB127A2 CRC32;
Query Match 81.3%; Score 39; DB 1; Length 187;
Best Local Similarity 83.3%; Pred. No. 2.59e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 151 EDKISY 156
:|||||
QY 1 ENKISY 6

RESULT 12
ID CYB_NOTPE STANDARD; PRT; 381 AA.
AC O03538;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME B.
DE MTCTB OR COB OR CYTB.
OS NOTHOPROCTA PERDICARIA (CHILEAN TINAMOU).
OG MITOCHONDRION.

```

```

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC PALAEOGNATHAE; TINAMIFORMES; TINAMIDAE; NOTHOPROCTA.
[1]
RN SEQUENCE FROM N.A.
RA LEE K., FEINSTEIN J., CRACRAFT J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";
RL (IN) MINDELL D.P. (EDS.);
RL AVIAN MOLECULAR EVOLUTION AND SYSTEMATICS, PP.1-1, ACADEMIC PRESS,
RL NEW YORK (1997).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q +
CC 2 FERROCYTOCHROME C.
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U76053; G2198708; -.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; 1.
DR PFAM; PF00032; cytochrome_b_C; 1.
DR PFAM; PF00033; cytochrome_b_N; 1.
DR ELECTRON TRANSPORT; MITOCHONDRION; RESPIRATORY CHAIN; TRANSMEMBRANE;
KW HEME.
KW METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 381 AA; 42895 MW; 71AC7B13 CRC32;
Query Match 81.3%; Score 39; DB 1; Length 381;
Best Local Similarity 85.7%; Pred. No. 2.59e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 374 ENKILYQ 380
:|||||
QY 1 ENKISYQ 7

RESULT 13
ID PYRP_BACSU STANDARD; PRT; 434 AA.
AC P39766; P25982;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE URACIL PERMEASE (URACIL TRANSPORTER).
GN PYRP.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
[1]
RN SEQUENCE FROM N.A.
RA QUINN C.L., STEPHENSON B.T., SWITZER R.L.;
RT "Functional organization and nucleotide sequence of the Bacillus
RT subtilis pyrimidine biosynthetic operon.";
RL J. BIOL. CHEM. 266:9113-9127(1991).
[2]
RN REVISIONS, AND CHARACTERIZATION.
RP MEDLINE; 94266724.
RX TURNER R.J., LU Y., SWITZER R.L.;

```

RT Regulation of the Bacillus subtilis pyrimidine biosynthetic (pyr)
RT gene cluster by an autogenous transcriptional attenuation
RT mechanism.";
RL J. BACTERIOL. 176:3708-3722(1994).
CC -!- FUNCTION: TRANSPORT OF URACIL IN THE CELL.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M59757; G387578; -;
CC EMBL; Z99112; E1185140; -;
CC SUBTILIST; BG10992; PYRP.
CC PROSITE; PS01116; XANTH_URACIL_PERMASE; 1.
CC PFAM; PF00860; xan_ur_permease; 1.
CC TRANSPORT; TRANSMEMBRANE.
FT TRANSMEM 17 37 POTENTIAL.
FT TRANSMEM 46 62 POTENTIAL.
FT TRANSMEM 69 85 POTENTIAL.
FT TRANSMEM 92 111 POTENTIAL.
FT TRANSMEM 118 140 POTENTIAL.
FT TRANSMEM 164 181 POTENTIAL.
FT TRANSMEM 188 207 POTENTIAL.
FT TRANSMEM 276 292 POTENTIAL.
FT TRANSMEM 301 317 POTENTIAL.
FT TRANSMEM 324 340 POTENTIAL.
FT TRANSMEM 347 364 POTENTIAL.
FT TRANSMEM 376 392 POTENTIAL.
FT TRANSMEM 401 418 POTENTIAL.
SQ SEQUENCE 434 AA; 45572 MW; F5EFF661 CRC32;

Query Match 81.3%; Score 39; DB 1; Length 434;
Best Local Similarity 57.1%; Pred. No. 2.59e+01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 365 DNKIDYE 371
QY 1 ENKISYQ 7

RESULT 14
ID YGER_ECOLI STANDARD; PRT; 458 AA.
AC P76639; Q45788;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 52.8 KD PROTEIN IN KDUI-LYSS INTERGENIC REGION.
GN YGER.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., ROSE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
CC -!- SIMILARITY: TO SALMONELLA INVASION PROTEIN IAGA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; U28375; G887803; -;
DR EMBL; AE000369; G1789216; -;
DR ECOCENE; EG13038; YGEH.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 458 AA; 52752 MW; 4ECE662D CRC32;

Query Match 81.3%; Score 39; DB 1; Length 458;
Best Local Similarity 71.4%; Pred. No. 2.59e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 4 ENKFSYH 10
QY 1 ENKISYQ 7

RESULT 15
ID TH13 YEAST STANDARD; PRT; 609 AA.
AC Q07471; P89098;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THIAMINE METABOLISM REGULATORY PROTEIN TH13.
GN TH13 OR YDL080C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE OF 42-609 FROM N.A.
RC STRAIN-GRF88;
RA NISHIMURA H., NOSAKA K., KANEKO Y., WATANABE K., IWASHIMA A.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.

Query Match 81.3%; Score 39; DB 1; Length 609;
Best Local Similarity 71.4%; Pred. No. 2.59e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 350 ESKLSYQ 356
QY 1 ENKISYQ 7

RESULT 15
ID TH13 YEAST STANDARD; PRT; 609 AA.
AC Q07471; P89098;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THIAMINE METABOLISM REGULATORY PROTEIN TH13.
GN TH13 OR YDL080C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE OF 42-609 FROM N.A.
RC STRAIN-GRF88;
RA NISHIMURA H., NOSAKA K., KANEKO Y., WATANABE K., IWASHIMA A.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.

Query Match 81.3%; Score 39; DB 1; Length 609;
Best Local Similarity 71.4%; Pred. No. 2.59e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 350 ESKLSYQ 356
QY 1 ENKISYQ 7

Search completed: Thu Oct 28 11:34:50 1999
Job time : 7 secs.

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

Run on: Thu Oct 28 11:35:06 1999; MasPar time 4.26 Seconds

Tabular output not generated. 89.788 Million cell updates/sec

Title: >US-09-081-707-12

Description: (1-7) from US09081707.pcp

Perfect Score: 48

Sequence: 1 ENKISYQ 7

Scoring table: PAM 150

Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: sprenbl9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mnc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 20.335; Variance 21.516; scale 0.945

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	46	95.8	51	2	Q53465 6 KDA ANTIGEN (FRAGMEN	2.29e+00
2	46	95.8	546	2	Q51850 GS60 ANTIGEN	2.29e+00
3	45	93.8	287	2	O85359 28 KDA MAJOR SURFACE A	4.00e+00
4	42	87.5	875	5	O76898 T17A3.1 PROTEIN.	2.04e+01
5	42	87.5	879	2	P75377 F11_ORF879 PROTEIN.	2.04e+01
6	42	87.5	1021	2	O25200 HYPOTHETICAL 120.1 KD	2.04e+01
7	41	85.4	282	10	O65412 HYPOTHETICAL 31.5 KD P	3.44e+01
8	41	85.4	292	2	O54429 PLASMID PIL2614 REPLIC	3.44e+01
9	41	85.4	312	2	Q47526 REVERSE TRANSCRIPTASE.	3.44e+01
10	41	85.4	544	1	P95960 PHENYLALANYL-TRNA SYNT	3.44e+01
11	41	85.4	704	6	Q29545 PORCINE INHIBITOR OF C	3.44e+01
12	40	83.3	232	11	O61538 ENDOGENOUS MAMMARY TUM	5.75e+01
13	40	83.3	362	11	O61536 ENVELOPE POLYPEPTIDE (5.75e+01
14	40	83.3	374	2	P73953 HYPOTHETICAL 39.6 KD P	5.75e+01
15	40	83.3	462	5	O44316 F56A11.3 PROTEIN.	5.75e+01
16	40	83.3	530	2	O87219 TRSK PROTEIN (TRAK).	5.75e+01
17	40	83.3	612	5	O44048 ASPARAGINE-RICH PROTEI	5.75e+01
18	40	83.3	967	2	O26050 TYPE III RESTRICTION E	5.75e+01
19	40	83.3	1888	11	O88466 ZINC FINGER PROTEIN 10	5.75e+01
20	40	83.3	2958	3	Q12150 CHROMOSOME XII COSMID	5.75e+01

21 39 81.3 144 2 Q03635 AMINOGLYCOSIDASE-6'-AC 9.53e+01
22 39 81.3 287 1 O26425 HYPOTHETICAL 32.4 KD P 9.53e+01
23 39 81.3 289 14 O92465 HE65-ACMNPV ORF105. 9.53e+01
24 39 81.3 344 2 Q99003 ADHESIN PRECURSOR. 9.53e+01
25 39 81.3 350 2 P94490 XYLOSE REPRESSOR. 9.53e+01
26 39 81.3 378 3 Q12076 SAGA. 9.53e+01
27 39 81.3 394 2 O25958 HYPOTHETICAL 45.1 KD P 9.53e+01
28 39 81.3 417 2 O25177 HYPOTHETICAL 47.5 KD P 9.53e+01
29 39 81.3 420 5 Q17468 B0284.4 PROTEIN. 9.53e+01
30 39 81.3 484 1 O58908 484AA LONG HYPOTHETICA 9.53e+01
31 39 81.3 593 5 Q23413 COSMID ZK1193. 9.53e+01
32 39 81.3 635 2 Q48484 O1:K20 RFBC (O1:K20 IN 9.53e+01
33 39 81.3 664 6 Q02849 PEPTIDYLARGININE DEIMI 9.53e+01
34 39 81.3 692 5 Q17325 CELFURPC PROTEIN. 9.53e+01
35 39 81.3 692 5 Q17797 FLA6.1A PROTEIN. 9.53e+01
36 39 81.3 747 13 Q91900 COMPLEMENT FACTOR B. 9.53e+01
37 39 81.3 760 5 Q17798 FLA6.1B PROTEIN. 9.53e+01
38 39 81.3 1085 5 Q24363 I FACTOR. 9.53e+01
39 39 81.3 1234 5 Q24690 HYPOTHETICAL 141.0 KD 9.53e+01
40 39 81.3 1421 10 O80907 T19C21.7 PROTEIN. 9.53e+01
41 39 81.3 2058 5 Q25757 CTRP. 9.53e+01
42 39 81.3 2166 2 O51465 HYPOTHETICAL 254.2 KD 9.53e+01
43 39 81.3 2643 5 O01552 HYPOTHETICAL 294.4 KD 9.53e+01
44 38 79.2 318 13 Q91974 REL-ASSOCIATED PP40. 1.56e+02
45 38 79.2 535 5 O76913 EG:95B7.10 PROTEIN. 1.56e+02

ALIGNMENTS

RESULT 1
ID Q53465 PRELIMINARY; PRT; 51 AA.
AC Q53465;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 6 KDA ANTIGEN (FRAGMENT).
OS PASTEURILLA HAEMOLYTICA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURILLACEAE;
OC PASTEURILLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95027654.
RA WELDON S.K., MOSTER D.A., SIMONS K.R., CRAVEN R.C., CONFER A.W.;
RT "Identification of a potentially important antigen of Pasteurella haemolytica."
RL VET. MICROBIOL. 40:283-291(1994).
DR EMBL; S74144; E155751; .
FT NON_TER 1
FT NON_TER 51
SQ SEQUENCE 51 AA; 5954 MW; 93E7B40E CRC32;

Query Match 95.8%; Score 46; DB 2; Length 51;
Best Local Similarity 85.7%; Pred. No. 2.29e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 12 ENKISYQ 18
QY 1 ENKISYQ 7

RESULT 2
ID Q51850 PRELIMINARY; PRT; 546 AA.
AC Q51850;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE GS60 ANTIGEN.
GN GS60.
OS PASTEURILLA HAEMOLYTICA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURILLACEAE;
OC PASTEURILLA.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-SEROTYPE 1;
RA LO R.Y.C., MELLORS A.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U42028; G1353671;
SQ SEQUENCE 546 AA; 60815 MW; 84B6A9DA CRC32;

Query Match 95.8%; Score 46; DB 2; Length 546;
Best Local Similarity 85.7%; Pred. No. 2.29e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 57 EDKISYQ 63
I:|||||
QY 1 ENKISYQ 7

RESULT 3
ID O85359 PRELIMINARY; PRT; 287 AA.
AC O85359;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 28 KDA MAJOR SURFACE ANTIGEN-1.
OS EHRlichia CANIS.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
OC RICKETTSIAEAE; EHRlichiae; EHRlichia.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-OKLAHOMA;
RX MEDLINE; 98321180.
RA REDDY G.R., SULSONA C.R., BARBET A.F., MAHAN S.M., BURRIDGE M.J.,
RA ALLEMAN A.R.;
RT "Molecular characterization of a 28 kda surface antigen gene family
of the tribe Ehrlichiae";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 247:636-643(1998).
DR EMBL; AF062762; G3327965;
SQ SEQUENCE 287 AA; 32014 MW; 1216502E CRC32;

Query Match 93.8%; Score 45; DB 2; Length 287;
Best Local Similarity 85.7%; Pred. No. 4.00e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 214 QNKISYQ 220
I:|||||
QY 1 ENKISYQ 7

RESULT 4
ID O76698 PRELIMINARY; PRT; 875 AA.
AC O76698;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE T17A3.1 PROTEIN.
GN T17A3.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COFSEY T., COOPER J., FAVELLO A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";

RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA CLARKE K., ROHLFING T., MORRIS M.;
RT "The sequence of C. elegans cosmid T17A3";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF078787; G3329644;
SQ SEQUENCE 875 AA; 99346 MW; 57A375A8 CRC32;

Query Match 87.5%; Score 42; DB 5; Length 875;
Best Local Similarity 85.7%; Pred. No. 2.04e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 528 ERKISYQ 534
I:|||||
QY 1 ENKISYQ 7

RESULT 5
ID P75377 PRELIMINARY; PRT; 879 AA.
AC P75377;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE FILORF879 PROTEIN.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae";
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-M129;
RA HIMMELREICH R., HILBERT H., LI B.-C.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE000042; G1674117;
SQ SEQUENCE 879 AA; 101086 MW; 94D217E0 CRC32;

Query Match 87.5%; Score 42; DB 2; Length 879;
Best Local Similarity 71.4%; Pred. No. 2.04e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 96 DNKIRYQ 102
I:|||||
QY 1 ENKISYQ 7

RESULT 6
ID O25200 PRELIMINARY; PRT; 1021 AA.
AC O25200;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 120.1 KD PROTEIN.
GN HP0453

OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-26695;
 RX MEDLINE: 97394467.
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
 RA FLETSCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATHEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.;
 RA "The complete genome sequence of the gastric pathogen *Helicobacter*
 RT *pylori* published erratum appears in Nature 1997 Sep
 RT 25; 389(6649):412.";
 RL NATURE 388:539-547(1997).
 DR EMBL: AE000560; G2313560; -
 DR TIGR: HP0453; -
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 1021 AA; 120112 MW; 3002FBB2 CRC32;
 Query Match 87.5%; Score 42; DB 2; Length 1021;
 Best Local Similarity 71.4%; Pred. No. 2.04e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 558 EKIAYQ 564
 QY 1 ENKISYQ 7
 RESULT 7
 ID O65412 PRELIMINARY; PRT; 282 AA.
 AC O65412;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 31.5 KD PROTEIN.
 GN F185.90.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., PETERS S.A., VAN STAVEREN M., DIRKSE W., STIEKEMA W.,
 RA BANCROFT I., MEWES H.W., MAYER K., SCHUELLER C.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL022603; E1287836; -
 KW HYPOTHETICAL PROTEIN
 SQ SEQUENCE 282 AA; 31535 MW; E5572824 CRC32;
 Query Match 85.4%; Score 41; DB 10; Length 282;
 Best Local Similarity 71.4%; Pred. No. 3.44e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 124 ESKISYH 130
 QY 1 ENKISYQ 7
 RESULT 8
 ID O54429 PRELIMINARY; PRT; 292 AA.
 AC O54429;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PLASMID PIL2614 REPLICATION PROTEIN
 DE (REPB), TYPE IC RESTRICTION SUBUNIT
 DE (HSDR), TYPE IC MODIFICATION SUBUNIT
 DE (HSDM), TYPE IC MODIFICATION SUBUNIT

DE (HSDS), AND PHAGE ABORTIVE INFECTION PROTEIN
 DE (ABI420) GENES, COMPLETE CDS (HSDR) (HSDS) (ABI420).
 OS LACTOCOCCUS LACTIS (STREPTOCOCCUS LACTIS).
 OG PLASMID PIL2614.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IL1403;
 RX MEDLINE: 98101482.
 RA SCHOULER C., CLIER F., LERAYER A.L., EHRLICH S.D., CHOPIN M.C.;
 RT "A type IC restriction-modification system in *Lactococcus lactis*.";
 RL J. BACTERIOL. 180:407-411(1998).
 DR EMBL: U90222; G2865241; -
 KW PLASMID.
 SQ SEQUENCE 292 AA; 34223 MW; 9EE797D CRC32;
 Query Match 85.4%; Score 41; DB 2; Length 292;
 Best Local Similarity 71.4%; Pred. No. 3.44e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 23 QNKISYE 29
 QY 1 ENKISYQ 7
 RESULT 9
 ID Q47526 PRELIMINARY; PRT; 312 AA.
 AC Q47526;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE REVERSE TRANSCRIPTASE.
 GN RET.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CLINICAL STRAIN;
 RX MEDLINE: 93116591.
 RA LIM D.;
 RT "Structure and biosynthesis of unbranched multicopy single-stranded
 RT DNA by reverse transcriptase in a clinical *Escherichia coli*
 RT isolate.";
 RL MOL. MICROBIOL. 6:3531-3542(1992).
 DR EMBL: Z12832; G42501; -
 DR PFAM: PF00078; rvt; 1.
 KW RNA-DIRECTED DNA POLYMERASE.
 SQ SEQUENCE 312 AA; 35734 MW; E0E2B0BE CRC32;
 Query Match 85.4%; Score 41; DB 2; Length 312;
 Best Local Similarity 100.0%; Pred. No. 3.44e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 175 ENKISY 180
 QY 1 ENKISY 6
 RESULT 10
 ID P95960 PRELIMINARY; PRT; 544 AA.
 AC P95960;
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE PHENYLALANYL-TRNA SYNTHETASE ALPHA CHAIN.
 OS SULFOLOBUS SOLFATARICUS.
 OC ARCHAEA; CRENARCHAEOTA; SULFOLOBALES; SULFOLOBUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-P2;
 RX MEDLINE: 97055432.

RA SENSEN C.W., KLENK H.P., SINGH R.K., ALLARD G., CHAN C.C.Y., LIU Q.Y.,
RA PENNY S.D., YOUNG F., SCHENK M.E., GAASTERLAND T., DOOLITTLE W.F.,
RA RAGAN M.A., CHARLEBOIS R.L.;
RT "Organizational characteristics and information content of an
archaeal genome: 156 kb of sequence from Sulfolobus solfataricus
P2";
RT MOL. MICROBIOL. 22:175-191(1996).
RL EMBL: Y08257; E283832; -.
KW AMINOACYL-TRNA SYNTHETASE.
SQ SEQUENCE 544 AA; 61815 MW; 1C2FFA45 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 544;
Best Local Similarity 57.1%; Pred. No. 3.44e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 459 DNKYSY 465
:|||||
QY 1 ENKISY 7

RESULT 11
ID Q29545 PRELIMINARY; PRT; 704 AA.
AC Q29545;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PORCINE INHIBITOR OF CARBONIC ANHYDRASE PRECURSOR.
GN PICA.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93099129.
RA ROUSH E.D., FIERKE C.A.;
RT "Purification and characterization of a carbonic anhydrase II
inhibitor from porcine plasma";
RL BIOCHEMISTRY 31:12536-12542(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97254619.
RA WUEBBENS M.W., ROUSH E.D., DECASTRO C.M., FIERKE C.A.;
RT "Cloning, sequencing, and recombinant expression of the porcine
inhibitor of carbonic anhydrase: a novel member of the transferrin
family";
RT family";
RL BIOCHEMISTRY 36:4327-4336(1997).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
OF AN ANION, USUALLY BICARBONATE.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
DR EMBL: U36916; G1016330; -.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 1.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
DR PFAM; PF00405; transferrin; 2.
KW SIGNAL; IRON TRANSPORT; GLYCOPROTEIN; METAL-BINDING.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 704 PORCINE INHIBITOR OF CARBONIC ANHYDRASE.
SQ SEQUENCE 704 AA; 77634 MW; 2E961A99 CRC32;

Query Match 85.4%; Score 41; DB 6; Length 704;
Best Local Similarity 57.1%; Pred. No. 3.44e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 667 QDKITY 673
:|||||
QY 1 ENKISY 7

RESULT 12
ID Q61538 PRELIMINARY; PRT; 232 AA.
AC Q61538;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENDOGENOUS MAMMARY TUMOR VIRUS (MMTV) RNA, ENV GENE AND RIGHT LTR
(MMTV).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85033926.
RA KWON B.S., WEISSMAN S.M.;
RT "Mouse mammary tumor virus-related sequences in mouse lymphocytes are
inducible by 12-O-tetradecanoyl phorbol-13-acetate";
RL J. VIROL. 52:1000-1004(1984).
DR EMBL: M11024; E19127; -.
SQ SEQUENCE 232 AA; 25722 MW; 697C4E0D CRC32;

Query Match 83.3%; Score 40; DB 11; Length 232;
Best Local Similarity 83.3%; Pred. No. 5.75e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 119 DNKISY 124
:|||||
QY 1 ENKISY 6

RESULT 13
ID Q61536 PRELIMINARY; PRT; 362 AA.
AC Q61536;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE POLYPROTEIN (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85033926.
RA KWON B.S., WEISSMAN S.M.;
RT "Mouse mammary tumor virus-related sequences in mouse lymphocytes are
inducible by 12-O-tetradecanoyl phorbol-13-acetate";
RL J. VIROL. 52:1000-1004(1984).
DR EMBL: M11024; G193081; -.
KW POLYPROTEIN; ENVELOPE PROTEIN.
FT NON_TER 1
SQ SEQUENCE 362 AA; 40429 MW; DD234921 CRC32;

Query Match 83.3%; Score 40; DB 11; Length 362;
Best Local Similarity 83.3%; Pred. No. 5.75e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 249 DNKISY 254
:|||||
QY 1 ENKISY 6

RESULT 14
ID P73953 PRELIMINARY; PRT; 374 AA.
AC P73953;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 39.6 KD PROTEIN.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.

Search completed: Thu Oct 28 11:35:53 1999
Job time : 47 secs.

RC STRAIN-PCC6803;
RX MEDLINE: 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIRASAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT genome and assignment of potential protein-coding regions.";
RL DNA RES. 3:109-136(1996).
DR EMBL: D90911; D1018753; -;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 374 AA; 39647 MW; F80260FC CRC32;

Query Match 83.3%; Score 40; DB 2; Length 374;
Best Local Similarity 71.4%; Pred. No. 5.75e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 127 ESKISYE 133
QY 1 ENKISYQ 7

RESULT 15
ID O44516 PRELIMINARY; PRT; 462 AA.
AC O44516;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE F56A11.3 PROTEIN.
DE F56A11.3.
GN F56A11.3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX GATTUNG S., GOELA D., HARPER M.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF038619; G2702456; -;
SQ SEQUENCE 462 AA; 52648 MW; 7FB30B7A CRC32;

Query Match 83.3%; Score 40; DB 5; Length 462;
Best Local Similarity 57.1%; Pred. No. 5.75e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 372 DERISYQ 378
QY 1 ENKISYQ 7

THIS PAGE BLANK (USPTO)